

Towards the patenting of animals

The decision in the United States that engineered strains of animals are covered by patent law has begun a debate among European patents-watchers that must lead to an amendment of European law.

THE staff of *Nature* will be among the first people in the world to be able to grow a new variety of rose named 'Harold Macmillan' after the former British prime minister whose family publishing company founded and still owns this magazine. (The US Macmillan Publishing Company in the United States, just acquired by Mr Robert Maxwell's companies, was established by the Macmillan family, which sold it, together with the right to trade in the United States under the family name, in response to the British government's requirement that British companies should sell US assets to pay for wartime Lend-Lease of military equipment.) While the breeder of the Harold Macmillan rose will retain exclusive rights to sell the variety, there will be nothing to prevent green-fingered *Nature* staff from propagating the rose or using it as a stock from which to derive a new variety. By contrast, those who buy the 'oncomouse' on which a Harvard University group has been awarded a patent, and which is the first patented animal to be offered for sale (see page 300), will be much more tightly restricted in the uses they can make of it.

Why should there be such a disparity? In the United States as elsewhere, there is a tradition that plant varieties can be registered (giving breeders the sole right to sell them) but not patented, but there are no precedents on animals. But the latest developments on this front in the United States have triggered great anxiety and debate in Europe.

One catalyst of the debate is biotechnology, by means of which mere molecular biologists can produce plants and animals with new genes in a fraction of the time taken to produce a new variety traditionally. The 'oncomouse' is an example, the result of experimentally introducing a cancer-promoting oncogene into mouse eggs so that the gene is carried by the resulting 'transgenic' mice and can be transferred to subsequent generations. It has taken only four years to produce, patent and, now, sell the first 'oncomouse'. But the European Patent Office has so far not made up its mind about 'oncomouse'; if, as expected, it rejects the claim, that may simply mean that it wishes to force arbitration by an appeal board.

'1992' is another catalyst of debate. That is the year in which trade barriers between the member states of the European Community will be eliminated. In anticipation, and to make it easier for companies to operate within a single commercial environment, there is a rush to standardize all manner of laws and regulations, among them the patent law for biotechnology.

Patent law in Britain, as in most European countries, specifically excludes the patenting of plant or animal varieties, while allowing patenting of microorganisms, which have come to include such modern entities as hybrid cell lines from which monoclonal antibodies are produced. One reason for the bias against higher organisms in the British Patents Act (1977) is that there is established and seemingly satisfactory legislation which provides plant breeders with rights over new varieties they produce. But the rights are limited, which is why those who grow the Harold Macmillan rose will be able not only freely to propagate and derive new varieties, but also to sell roses from the plant without paying a royalty.

But can this permissive system last? Difficulties had arisen even before the advent of gene technology, but have now been

sharpened, especially in relation to the plant equivalents of transgenic animals. Can it be appropriate to apply the existing plant breeders' legislation which, like its equivalents throughout the world, was drawn up when plants that are now a reality, such as tobacco endowed with a gene that makes it resistant to a particular herbicide, were at best pipe-dreams?

The answer has to be no, not only because the means and genes involved can themselves be patented, but because a plant with a single new gene can be brought to the market and further engineered genetically, and more rapidly than by traditional breeding. And while it must be next to impossible faithfully to repeat the process by which a new variety is produced by traditional breeding, the addition of a defined gene can more easily be copied. Patent protection must therefore become the norm for transgenic plants, and fiscal systems will have to be devised whereby those who produce the new varieties can share the benefits their users will enjoy. So much is recognized in the draft proposal for a new European patent law for biotechnology issued by the European Commission last month.

That proposal also recognizes the need to provide patent protection for animals. Matters are simplified because there is no present system of protection of varieties to be taken into account. Whether, as some argue, there needs to be an equivalent of plant breeders' rights as well as patent protection will largely depend on the ingenuity of those who draft the new laws.

Where does this leave the practitioner molecular biologists, many of whom have retained their innocence? The jungle of existing and evolving patent law is often impenetrable, but some will have to be made to grasp the basics. But is there any more reason for the details of patent law to be familiar to practitioners than for the details of Southern blotting to patent lawyers? Biotechnology companies are probably able to provide the legal help their people need, but are research institutes and universities ready to meet what could be a taxing challenge? □

University nannies

The British government must decide whether its autonomous universities should be autonomous.

JUST what the British government intends for British universities is increasingly a matter for conjecture. Less than six months have gone by since the British Parliament was locked in battle over Mr Kenneth Baker's Education Reform Bill, now safely on the statute book. Among the several famous battles waged on that occasion were that in which the government agreed to accept that academic freedom is a concept that properly belongs in universities and that is not so vague that it cannot be enshrined in the English language. Among other things, the government also undertook that, in its new regime for the British universities, there would be an understanding that universities successfully following its exhortations to raise more money from private sources would no longer be penalized in the annual share-out of government subventions. Thus British law now includes the provision that the Universities Funding Council (the first word lacks an apostrophe despite the government's preoccupation



with the teaching of grammar) shall take no account of the volume of private support a university enjoys when deciding what its support from the public purse will be.

But somewhere, sadly, people seem to be having second thoughts. The Universities Funding Council (UFC) is now a going concern, with the head of the Cranfield Institute of Technology as its chairman. Although UFC will formally take over from its University Grants Committee (UGC) only on 1 April next, the beavers appear to be hard at work determining the rules by which it will operate. Some of these have been leaked to the press in the past few days, and one of them is the requirement that UFC will be entitled to enquire into the amount of a university's reserves.

This is a curious development, and has understandably incensed many university managers for reasons which the government appears unable to understand. For what the universities say, quite simply, is that there is nothing in their present relationship with UGC that sanctions such inquisitiveness, and the upshot of the debate about the education bill was that the government would retreat from the appearance of its first intentions to manage the universities from the centre as if they were so many branches in a chain of department stores. But how, the government seems to reply, can UFC exercise proper care of its dependants, both the whole flock of its charges and the individual chickens within it, unless it has some means of telling whether all of them are in good health? And what better way of doing that than by monitoring the private books?

The dangers, of course, are obvious. While UFC will be enjoined not to be prejudiced against universities successful in raising private funds, only exceptionally strong-minded people will be able to follow that precept in deciding what funds should be allocated to universities with ample reserves and those with none, especially when the going is rough. And while prejudice of this kind will be against the law as it now stands, there will be no way of proving that. UFC might simply say that it is not prejudiced against private funds, but merely against institutions that are well-off for whatever reason. In any case, nobody would be sent to gaol, for no penalties are specified. And the universities' indignation is probably misplaced, for even if UFC neglected to ask about their reserves, charity law would probably require that their value should be published once a year.

So is the row just another storm in a tea-cup? It would be good if it were so. But the suggestion that UFC should have such detailed powers of inspection over universities points to an inconsistency at the heart of its brave new policy. At the outset, when the Education Reform Bill was published, the government seemed bent on intervention, direction from the centre. With the passage of the bill, it seemed to have learned that there is merit in a system in which universities are made autonomous and are impelled towards the responsible conduct of their own affairs by the knowledge that, if they do otherwise, they will become second-rate, even third-rate, or may even go under altogether. In the long run, that is the way towards a greater diversity within the British system of higher education which, in turn, is what the British national interest requires. And that is why the caring supervisory role now being thrust on UFC by the government is at odds with what should be public policy. What is it about universities that makes British civil servants believe that they could run them more efficiently than those chosen by academics to do so?

Strangely, the government has also managed to offend its other constituency in higher education, the British polytechnics that will be detached from local authority control next April, and subject instead to an analogous funding council. Ironically, since the publication of the education bill over which the universities made such a fuss, the polytechnics have been generally more compliant, saying how much they welcomed the opportunity apparently on offer to run themselves in future as if they were businesses happening to be in the business of higher education. So why should the draft terms of the government's

instruction to the polytechnics' funding council lay down that institutions selling property they happen to own on 1 April will not necessarily be able to keep the whole of the proceeds for themselves? The obvious danger is that some of the city-centre polytechnics might be able powerfully to maximize their resources by selling off some of their sites and moving elsewhere. But would it necessarily be a bad thing that fuller use should be made of economic resources in this way? Or that some polytechnics might be able so to better themselves in the process that the quality of higher education would be significantly improved? To believe otherwise would be inconsistent with the government's whole philosophy. □

Uncensored Nature

A canard that Nature has withheld information about fallout from Chernobyl is untrue.

SOME Finnish newspapers last week carried a strange allegation to the effect that *Nature* has held up publication of data gathered in the weeks after the accident at Chernobyl on 6 April 1986, allegedly because of self-censorship. One version of the tale is that data have been suppressed for fear of giving offence to the British government, although the logic of that assertion is hard to credit: in the months following the accident, the British government seemed to this journal to be all too welcoming of alarming news, and all too ready to clap restrictions on sheep farmers. Would it not have been sustained by the publication of even more data, however scary?

The truth, sadly, is more mundane. In the weeks following the accident at Chernobyl, a great many notes were received reporting then recent measurements of fallout from the accident, and were quickly published without being refereed in the usual way. Predictably, the volume of correspondence quickly grew, and with it the prospect that other branches of science would be swamped by fallout measurements. There came a time, in the second half of 1986, when it seemed prudent to call a halt, but there were a large number of notes submitted for publication between the time the shutters came down and that at which the flow of intended publications tailed off. During this hiatus, some contributors were told that their notes would be published (for which the Editor takes personal responsibility). Those have been an embarrassment to *Nature* and to their authors ever since.

The obvious difficulty is that, with the passage of time, the information immediately available has been overtaken by more studied measurements and more careful analyses of them. At one stage, it seemed possible that these notes might be lent freshness by the publication of a symposium to mark the first anniversary of the accident, but that came to nothing, most probably because there had been too little time. Earlier this year, the Editor wrote to the, by then, frustrated contributors, canvassing their wishes but suggesting that the unpublished material would at that stage serve little purpose. Rather more than half, with more or less reluctance, agreed. The remainder, on a point of principle, think otherwise; in all probability, last week's report in the Finnish newspapers reflects the puzzlement of such a source. But the notion that *Nature* might be susceptible to censorship is the strangest fallout from Chernobyl.

The issue is all the more complicated because there remains much to say about the fallout from the Chernobyl accident, especially from the Soviet side. Elsewhere, there remains a great deal to be done in working towards a better understanding of the relationship between radiological exposure and human health. *Nature*, which has no wish to fall out with any part of the scientific community, has particular need of the assistance of the radiological community just now, and therefore apologizes to those members of it who are aggrieved. In return, it would ask that they do not give currency to a canard which is self evidently untrue. □

US congressman attacks NIH investigation

■ Investigators accused of leaks ■ Misstatements admitted

Washington

THE National Institute's of Health (NIH) investigation of possible scientific misconduct in the 1986 paper in the journal *Cell* has been attacked by Representative John Dingell, the powerful Michigan Democrat who has been aggressively investigating the case. In a letter to Secretary of Health and Human Services Otis Bowen, Dingell claims that NIH leaked the conclusions of its investigation to Massachusetts Institute of Technology Nobel-laureate David Baltimore, one of the paper's authors, and colluded with Baltimore to allow him to preempt its findings.

In a 23-line note published in the 18 November issue of *Cell* (55, 541; 1988) Baltimore and three of the five co-authors of the disputed paper acknowledge that the paper contained "three instances of misstatement". But they say the corrections "are not material alterations and do not affect the conclusions of the paper".

The validity of the original paper (*Cell* 45, 247; 1986), was questioned in 1986 by Margot O'Toole, a postdoctoral researcher working in the laboratory of Thereza Imanishi-Kari, the paper's principal author. The controversy became public after Walter Stewart and Ned Feder, two NIH scientists known for their controversial inquiries into scientific misconduct, tried unsuccessfully to publish an analysis of the *Cell* paper (see *Nature* 332, 670; 1988) based partly on records alleged to have come from a notebook in Imanishi-Kari's laboratory.

What began as a minor scientific dispute soon turned into a major political issue. NIH began their own inquiry but ran into difficulties and had to begin again. Upset with the lack of progress in the inquiry, Dingell's powerful investigation and oversight subcommittee conducted a hearing on scientific misconduct that centred on the case. Since that hearing, there have been no public pronouncements on the case until last week's letter in *Cell*. NIH continued their own investigation, and a final draft was circulated last week.

In his account of a meeting between Baltimore, his attorney and the NIH panel earlier this autumn, Dingell claims his subcommittee staff was told by NIH general counsel Robert Lanman that Baltimore "was still not prepared to admit voluntarily that the *Cell* paper contained any errors". Dingell claims that Baltimore was able to infer the NIH panel's findings from the questions he was asked during

the meeting, "and realized what the panel would conclude". Dingell claims that Baltimore and his attorneys then asked what they could do, and "it was suggested . . . that they write the letter to *Cell* admitting the errors and misstatements".

Dingell goes on to accuse Baltimore and his co-authors of writing a letter to *Cell* that makes it appear that the errors in the original work had only recently come to their attention. "It is quite clear that this is not what actually happened: the authors were informed of these same misrepresentations two and a half years ago, but were only willing to admit them publicly when they realized they could no longer be kept secret."

Dingell's letter was prompted by a meeting on 4 November between attorneys for Baltimore and his own subcommittee's staff at which Baltimore's lawyers presented his note to *Cell*. They argued that as the three "misstatements" had been corrected, there was no need for Dingell's committee to continue its inquiry.

But Stewart is far from satisfied that the Baltimore note answers "the simple scientific arguments" he and Feder have made. The original paper presented evidence that the presence of a rearranged exogenous immunoglobulin gene affected the expression of the host mouse's own immunological repertoire. One important part of the data was provided by monoclonal antibodies that Baltimore and co-workers claimed would distinguish between the products of host and transplanted gene. In last week's note to *Cell*, they say they had made "an overstatement of its [the monoclonal antibody's] specificity".

While Baltimore now appears to concur with one of Stewart's main criticisms of the paper, Stewart comments that "they say it's an 'overstatement' but they don't say what the facts are or whether the specificity was adequate to allow them to do the job they claimed they had done".

The NIH inquiry seems likely to agree, at least in part. Although the report has not been officially released, sources say that it does not accept that Baltimore has answered all the doubts about his paper in his note to *Cell*.

Dingell is seeking an investigation by the inspector general of the Department of Health and Human Services of NIH's handling of the affair, and wants the inspector general to brief his subcommittee on his findings by 28 November.

Alun Anderson & Joseph Palca

Universities' heads accuse governments of interfering

London

RELATIONS between British universities and the government have taken a turn for the worse as the government attempts to draw up the guidelines for the relationship between itself, the new Universities Funding Council and the universities. A draft memorandum of the guidelines has been sent for consultation to the Committee of Vice-Chancellors and Principals (CVCP). The vice-chancellors claim that the guidelines would negate everything that was gained in universities' struggle for independence during the passage through parliament of the Education Reform Act 1988. The guidelines give the government a "licence to interfere" with the running of the universities, says Professor John Ashworth of the University of Salford. But the government says that such fears are groundless.

The vice-chancellors object to two clauses in the guidelines. One states that the university must seek the approval of the Secretary of State for Education before it can borrow money. The Department of Education says that approval will be necessary only in cases where the universities borrow money against public assets. But this would override assurances by the government that it would not interfere with individual institutions, leaving that to the funding council.

The other clause which has angered the vice-chancellors is one which says that the council will monitor universities' finances, making no distinction between public and private funds.

But the vice-chancellors had previously succeeded in adding to the Education Reform Bill before it became law an amendment which stated that the powers of the new funding council would apply only to funds from the council. The CVCP is now seeking legal advice as to whether this clause in the guidelines would be illegal and is confident that it would be.

The Department of Education says the concern of the CVCP over this clause is unnecessary because there exists in the act a provision that funds from public sources cannot be reduced simply because the university's income from private sources is increased.

But the CVCP claims that ministers are trying to claw back powers taken from the before the education Bill become law. And it is now seeking meetings with government ministers in order to hold the "meaningful discussions" which were promised by the Education Secretary, Mr Kenneth Baker, during the passage of the act.

Christine McGourty



A worker at the National Radio Observatory photographs the debris.

Radiotelescope in state of collapse

Washington

THE cause of the sudden collapse last week of the 300-foot-diameter radio telescope dish at Green Bank, West Virginia, remains a mystery. The radiotelescope, which is one of the world's largest and has been in use for 26 years, came crashing down last Tuesday night, bringing dozens of research projects to a halt.

According to Paul A. Vanden Bout, director of the National Radio Astronomy Observatory facility, there was "no wind, rain or snow" and the telescope had been working perfectly until the accident.

It is not even possible to tell whether the initial fault developed in the axle on which the dish is mounted or the two pylons holding the axle aloft, but all that is left is a heap of twisted metal. Nobody was hurt in the collapse, although the telescope operator in the control room was "shaken" and had considerable difficulty in convincing operators at Green Bank's other facility, a 140-foot-diameter radiotelescope, that the accident really had happened.

An internal inquiry team has begun an investigation, said Vanden Bout, but it is under instructions "not to disturb the remains in any way" until an external team

being hastily assembled by the National Science Foundation (NSF) arrives. Vanden Bout dismissed as "crazy" rumours that the 300-foot dish had been improperly designed.

The big dish was used for large-area sky searches and had played a key role in clarifying the origin of pulsars through the discovery of a pulsar at the heart of the Crab Nebula. At the time of collapse, the telescope was being used for a 6-cm radio sky catalogue. A first pass had been completed and a second pass, looking for sources that had changed, was three-quarters complete. Almost all the data will still be usable.

Telescopes at Arecibo, Puerto Rico, and Effelsberg, West Germany, provide some of the same northern sky coverage as the Green Bank telescope but neither matches its capacity exactly.

Rebuilding would be very expensive. Pat Bautz, NSF's director of the division of astronomical sciences, says the telescope cost \$850,000 in 1962. But astronomy is now very short of funds. Her first concern is to find out what went wrong.

Attempts are being made to reach NSF director Erich Bloch who is on his way to the Antarctic.

Alun Anderson

Expansion squeezed

Berkeley

BUDGET limitations passed by California voters may prevent the University of California (UC) from expanding as it struggles to keep up with the state's growing population.

To maintain its policy of accepting the top 12.5 per cent of the state's high-school graduates, the UC system will need to increase its total enrolment from its present size of 154,000 graduate and undergraduate students to about 217,000 by the year 2005. UC officials say the extra students cannot be accommodated on the existing nine campuses without compromising academic quality. UC president David P. Gardner has recommended the expansion of several of the existing campuses, in addition to building three new campuses, at a cost of \$300 million each, to open in 1998, 1999 and 2000.

Gardner's proposal may be in vain, however, because of voter-approved spending limitations imposed on the state. The 1979 'Gann limit', named for its sponsor, self-styled tax-revolt leader Paul Gann, puts a cap on state spending. An initiative passed by voters in November worsens the financial situation for UC, by assuring public primary and secondary schools and community colleges of a fixed percentage of each year's state budget, while giving UC no comparable assurance.

UC has made it clear that it will limit its enrolment rather than compromise academic quality, and if the state wants the university to continue to meet the needs of a growing population, the legislature must take action to reverse the voter-approved initiatives and remove the financial squeeze.

Marcia Barinaga

Celltech maturation

London

AN end to the special relationship between the UK Medical Research Council (MRC) and the leading UK biotechnology company, Celltech Ltd, has been announced. The eight-year-old company owes much of its initial success to its exclusive first rights to MRC discoveries in biotechnology. Five years ago, the company's privileges were reduced by MRC, and now all that remains is an agreement that the council will keep the company informed of technology developments in exchange for information of the company's interests and plans.

Celltech, meanwhile, has notched up its second profitable year, but only just. With turnover increasing from £11.5 million to £16.6 million in 1987, profits fell by nearly half to an unimpressive £125,000, and only because a considerable increase in interest income offset extra costs involved in developing potential pharmaceutical products.

Peter Newmark

Cautious delight at Baikonur as Soviet shuttle touches down

Baikonur

BURAN, a Soviet reusable spaceship, made its maiden flight on the morning of 15 November, at the second attempt. The first, two weeks earlier, was automatically aborted because of a fault in the communications. But this time the entire launching equipment, the Energia booster and the reusable spaceship, performed smoothly.

Work on the Energia-Buran space system has been under way in the Soviet Union for roughly ten years. In its technical concepts, it differs greatly from the US shuttle. Having plumped for shuttle spacecraft, the United States almost completely abandoned one-shot rockets. It was supposed that, with a sufficiently large number of flights — up to a dozen a year — shuttles would be more effective. But experience has shown that it is difficult to maintain such a schedule and cost savings have not been achieved.

Alexander Dunayev, head of Glavkosmos, says that the efficiency of manned space-flight, and its rewards, are also current problems in the Soviet Union. But, he says, "by using existing technology, the expense can be repaid". One Soviet objective now, he says, is to develop the commercial aspect of space.

Dunayev also says that it is "disadvantageous — both economically and technically — to use only one vehicle, which is why we have built several, bearing in mind that they will be in service for many years". There is no intention to use them intensively, making 20–30 flights a year, for example.

According to Dunayev, Soviet engineers have never raised the question of abandoning one-shot booster rockets, which will continue to perform the main work of space transport. But "reusable vehicles will make it possible to carry out some specific tasks such as the delivery of specialists to orbit, the repair of equipment in space or its removal back to the Earth".

As a rule, Dunayev says, reusable vehicles will be used when it is necessary to carry extra cargo to a ship already launched. As things are, he says, the cost of launching a payload by the reusable system is tens or hundreds of times greater than by a one-shot rocket.

The take-off mass of Energia-Buran is 2,400 tonnes, of which 90 per cent is fuel. The ship itself weighs 105 tonnes, with a 30-tonne payload. (The return payload is 20 tonnes.) While the characteristics of the Buran and the US shuttle are in many ways comparable, the Soviet system is more powerful — it has to be, however, because the US launching sites are nearer

the Equator.

Heating of Buran on return to the atmosphere can raise the external temperature to 1,600 °C, but the hull is protected by 30,000 quartz-carbon composite tiles with a total weight of 9 tonnes. Apart from the propulsion engines, there are two sets of smaller control engines near the crew cabin and in the tail section, all fed from a tank containing 14 tonnes of fuel. The maximum flight span is 30 days.

The Soviet Union began the development of Buran, in parallel with the Energia rocket, in 1977–78. The designers

are Valentin Glushko, Yuri Semenov and Boris Gubanov, who also designed the complicated launching and landing facilities. Apart from the 5-km landing strip near here, there will be two emergency landing strips in the west and east of the Soviet Union. Up to three reusable spacecraft will be used in the future, each of them launched up to four times a year.

On the reason why Buran's first flight was unmanned, Dunayev explains that it is important "to astronautics and for aviation" to demonstrate this capacity, but also that the first flight will help determine when the first manned Soviet flights will take place. He says that the second flight "may also be unmanned".

Mikhail Chernyshov
Novosti

Rival shuttle judged a success in the West

Washington

WESTERN experts are now asking how the Soviet shuttle programme was able to advance so far so quickly. Until the shuttle was unveiled last month, it was not clear that the Soviet Union had mastered the technology needed to fly a shuttle or even accepted the arguments for manned cargo-carrying space flights.

Soviet space experts have frequently criticized the US space shuttle programme as unnecessary and have stressed the development of large unmanned rockets. But the apparent change in direction may be partly related to internal Soviet politics, according to John Pike, director for space policy at the Federation of American Scientists. The shuttle project, he says, appears to have come out of the Aviation Ministry rather than the Machine Building Ministry which manufactures rockets and has dominated the space programme in the past.

A desire to keep up with the United States in shuttle design may be a factor, but it seems that the Soviet planners have come round to the view that a shuttle is the most convenient means of servicing and supplying orbiting manned space stations.

With Buran and its sister craft, the Soviet Union will be able to bring back scientific instruments, astronomical telescopes and machinery from the space station for repair as well as materials manufactured under gravity-free conditions. It will also be able to match the US shuttle's proven ability to pick up and return malfunctioning satellites.

Two teams of designers arriving at similar conclusions . . . or are they, in some way, related?

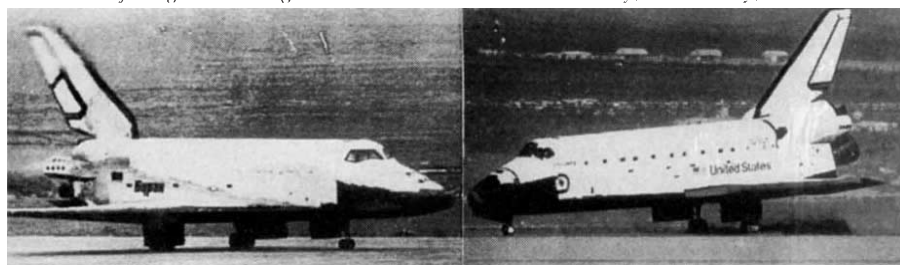
In the United States, there has been much speculation over how much Buran's designers learnt from studying the US space shuttle. But Pike claims that the "closer you look at the Soviet shuttle the less it looks like the US shuttle". He points out that the wing shape is different, the nose more pointed, and the payload bay doors narrower. Altogether he believes that "clear pictures would probably show more differences than between the Boeing 767 and the European Airbus".

Differences in design can be seen in such key items as the tiles which cover the shuttle and protect it from extreme heat as it re-enters the atmosphere. Each of the 22,000 tiles that cover the US shuttle is numbered and fits in only one place. But television close-ups of Buran show the tiles are not numbered and are apparently of a uniform design.

The biggest difference between the two shuttles, that Buran has no rocket engine of its own, may make the Soviet programme a more expensive one to run. Buran relies entirely on reaching orbit through the action of the enormous Energia rocket and four strap-on, liquid-fuel boosters. As the Energia rocket cannot be recovered, each launch costs the Soviet Union a new set of rocket motors. The US shuttle's three main engines are re-usable.

Soviet costs will also be pushed up because its overheads will be spread over fewer launches than the US programme.

Alun Anderson



Soviet Union to supply nuclear reactors to India

New Delhi

THE Soviet Union will build India's biggest nuclear power station, consisting of two 1,000-MW reactors, under an agreement signed on 20 November by Prime Minister Rajiv Gandhi and President Mikhail Gorbachev. A one-year pact on cooperation in exploration and use of outer space was also signed during Gorbachev's second visit to New Delhi.

The two Soviet VVER-type pressurized light water reactors will be located in the coastal town of Koodangulam in Tamil Nadu state. The reactors will be operational in the late 1990s and their import is the first by India since Canada and the United States severed nuclear ties in the aftermath of the Pokran nuclear test in 1974.

India has accepted the Soviet plants nearly eight years after the offer was first made by Leonid Brezhnev. India's own programme has been based on a Canadian design (CANDU) that uses locally available uranium and heavy water. The Soviet-built station will depend on imported enriched uranium fuel and newly trained operators.

The Soviet Union will build the reactors on a turnkey basis and supply fuel for their entire operational life. The spent fuel will be taken back by the Soviet Union. The Soviet Union will be committed to provide fuel irrespective of any future change in domestic laws — the United States stopped its fuel supply to the Tarapur reactors after the Indian nuclear test.

The Soviet-built atomic station will be under safeguards and will not be used for making nuclear weapons under a separate agreement with the International Atomic Energy Agency (IAEA). India had earlier accepted IAEA inspection in respect of the US-built Tarapur reactors and Canadian-built units near Kota in Rajasthan.

The addition of two more reactors to the 'safe-guarded' list is seen by some as erosion of India's nuclear self-reliance and sovereignty. The DAE secretary, Dr M. R. Srinivasan, has however said that the introduction of Soviet technology will have no adverse impact on his department's plans to build 25 or 30 CANDU reactors by the year 2000.

The long-term agreement on space signed by foreign ministers of both countries is an extension of the cooperation that began in 1963. It covers all areas in space science, space technology and applications. Specifically, there will be joint projects to study middle atmosphere and development of spaceborne microwave sensors. The Soviet Union will launch India's second remote sensing satellite IRS-IB in 1991.

K.S. Jayaraman

Harvard University attacked over ties with industry

Boston

AT a press conference at Harvard University last week, consumer activist Ralph Nader unveiled two reports documenting the extent of Harvard faculty members' ties to industry and accused the university of "incremental, relentless corporatization". The reports*, issued by a student-run, Nader-sponsored organization called 'Harvard Watch', disclose many instances of "conflicting commitments" by faculty members, and call upon the university to disclose publicly all corporate affiliations at the school.

According to the reports, Harvard faculty members are more involved in outside work with corporations and government than their peers at any other university in the United States. The reports claim, however, that these corporate affiliations are shrouded in secrecy, making information about them very difficult to obtain. Because of the lack of public scrutiny, conflicts of interest can often go undetected.

The reports cite the recent case of Scheffer Tseng, a researcher at a Harvard-affiliated hospital who, the university has acknowledged, was involved in a conflict of interest in clinical trials of an eye ointment for a company he helped to found (see *Nature* 335, 754; 1988). Harvard Watch criticized the university for knowing about Tseng's apparent conflict of interest but failing to disclose it publicly for at least five months before the case was first publicly exposed by the press.

Robert Weissman, the author of one of the Harvard Watch reports and an undergraduate senior at the university, stresses that the public expects disinterested expert advice from university professors. "We're not saying that university faculty shouldn't collaborate with industry", says Weissman, "but rather that students and the general public have a right to know about these relationships." Weissman points out that, in addition to faculty members' individual relationships with corporations, research contracts between the university and industry are not publicly disclosed. On the basis of limited financial information reported by Harvard, the Harvard Watch group asserts that such contracts exist with dozens of large corporate concerns including the Council for Tobacco Research, DuPont, Johnson and Johnson, R. J. Reynolds, Stroh Brewery and Union Carbide.

Nader expressed his concern that the university's principles of openness and availability of information are becoming subordinated to "mercantile values" dominated by the "rules of trade secrecy". Faculty commitments to corporations, he

said, rob students of the attention of faculty members, subvert the interests of research "subsidized by taxpayers" and threaten the "priceless heritage" of the university community. Nader says he hopes the reports on Harvard will "open the way for similar work at other universities". By reviewing the personnel on the boards of directors of the world's largest corporations, Harvard Watch has established that Harvard professors maintain 38 directorships with Fortune 500 companies, almost twice as many as are held by Massachusetts Institute of Technology faculty, the next nearest university competitor. In addition, the report cites specifically the dramatic increase in industry ties in the biological sciences over the past decade, with more than 60 faculty members now holding positions at biotechnology companies.

In a brief statement issued by the university in response to Nader's charges, Harvard spokesman Peter Costa stated that "relationships between universities and industry serve the public interest" and that the university is "committed to proceeding with such relationships thoughtfully". In its official statement, however, the university declined to comment on Harvard Watch's specific charges, in particular the call for public disclosure, claiming that the university has not had time to evaluate the charges and pointing to what they consider to be "extensive rules governing conflicts of interest or commitment by faculty members".

Harvard biologist and Nobel laureate Walter Gilbert says that corporate ties provide a number of benefits "including a variety of intellectual stimulation to faculty members". But Gilbert, who stressed that he had not seen the actual Harvard Watch reports, says he sees no reason why the university's or individual faculty members' corporate relationships should not be made public. "My own view is that any corporate-sponsored research should be open", he says. "Faculty members do have a number of reasons to collaborate with corporations, but there is no reason that those should be secret."

The greater danger, according to Gilbert is the university's involvement in a marketing venture, Ion, Inc. (see *Nature* 335, 291; 1988) which seeks to allow the university to profit from research at the medical school by bringing products to the market. Gilbert says that he thinks the university's reasons for establishing the venture are "foolish" and that it could lead to a "reduced ability to judge a conflict of interest". **Seth Shulman**

*Scholars, Inc.; Harvard Academics in Service of Industry and Government, and The New Classified Research: Corporate Sponsored Biomedical Research and the Reign of Secrecy at Harvard University, both from Harvard Watch, Cambridge, Massachusetts, November 1988.

UK government refuses to budge after Lords' attack on space policy

London

THE British government is often criticized for its attitude to space research, but last week it was the subject of a particularly scathing attack from a House of Lords committee. The Select Committee on Science and Technology said government policy on space research lacks inspiration and enthusiasm and that the British National Space Centre (BNSC) is ineffective. Responding to the attack, Lord Young, the Secretary for Trade and Industry, said that the committee's approach looks backward. He said that in asking for guidance and inspiration, the committee really wants greater public expenditure and a return to the days when government ministers tried to pick winners. He strongly disagreed that the BNSC is ineffective.

The criticisms followed the government's reply to the committee's report on national space policy, published in 1987. The reply was published last week. In it, the government ignored most of the Lords' criticisms, stressing that space projects would not be increased beyond the current level of £130 million a year. The Lords called for an increase to £200 million. But in his reply last week, Lord Young said that because space projects are long-term and "notoriously expensive", the government would be selective as regards the fields of activity it entered and would involve the private sector in selecting and developing future projects.

Salk president retires

Berkeley

FREDERIC de Hoffmann, president of the Salk Institute of Biological Studies in San Diego, announced his retirement last week, shortly after discovering he had been infected with the AIDS virus during a blood transfusion in 1984.

De Hoffmann, a physicist who was formerly president of Gulf General Atomic, became president of the Salk Institute in 1970. Known as a talented fund-raiser, he is credited not only with saving the institute from financial ruin, but with bringing in programmes in molecular biology and neuroscience, and engineering the institute's evolution from the visionary think-tank of its early years to a leader in experimental biology.

Rumours of his waning effectiveness that have circulated in recent years are now understood in light of his failing health.

Salk scientist and Nobel laureate Renato Dulbecco has been appointed acting president until a successor is found.

Marcia Barinaga

The government said that a major part of its civil space programme would continue to be carried out through international collaborations, particularly with the European Space Agency. But the committee retorted that unless Britain could summon up more enthusiasm for space it could not rely on its partners to continue collaborating.

The committee also said that it was disingenuous for the government to offer support to British Aerospace and Rolls-Royce to find suitable collaborators for the spaceplane HOTOL if it will provide no support itself. The costs of HOTOL and the timescale of commercial returns

are far beyond the scope of private industry and those companies are unlikely to secure foreign backing if the British government is unwilling to back the project, said the committee.

Criticizing the BNSC, which was set up to implement the country's civil space programme, the committee said it was "little more than an inter-departmental committee with a grander name" and that since the government will do nothing to strengthen BNSC, it obviously does not mind if the centre is ineffective. The committee had recommended that the BNSC be made independent with its own resources, that industrial partners be brought into it and that the director-general come from industry. But the government's response indicates that there will be no change.

Christine McGourty

Sakharov and Teller meet at last in Washington

Washington

A MEETING impossible to contemplate thirty years ago took place in a Washington hotel room last week between dissident Soviet physicist Andrei Sakharov, 67, 'father' of the Soviet hydrogen bomb, and Edward Teller, 80, credited with the same distinction for the US bomb.

The twenty-minute meeting took place during a photography session for a US

(SDI) to his friend President Ronald Reagan, emphasized that the technology of defence had not yet been given a sufficient trial, and that to abandon its promise would be a mistake.

Later, as Sakharov spoke at the dinner in Teller's honour, he noted how their lives had run along parallel tracks. Both men, he said, had been caught in a nationalistic fervour that made the development of a hydrogen bomb seem crucial. But now, Sakharov said, "I think that the work was a terrible tragedy". Before an audience enamoured of the promise of the SDI, Sakharov warned that "the world has entered a new era, and I am convinced that a new approach is necessary".

Teller, speaking after Sakharov had left for another engagement, agreed that their lives had moved along parallel lines. "Parallel, but of course different," Teller, a Hungarian Jew, was studying in Germany when the Nazis came to power. He moved to the United States in 1935 at the invitation of Russian expatriate George Gamow, then at George Washington University in Washington, DC.

Teller called the advances in technology the greatest hope for mankind, and noted that whereas he had been able to work in a laboratory and monitor the progress of the technology he helped to develop, his Soviet counterpart was forbidden access to sensitive information.

Neither man wavered in his commitment to principles that have guided their actions in recent years. But for one moment, as they sat and talked together, science and politics was put aside. "How is your heart", asked Teller of Sakharov who had seen a cardiologist in Boston. "It's OK", he replied, and they went down to dinner.

Joseph Palca



Teller and Sakharov last week.

news magazine just before a dinner sponsored by the Ethics and Public Policy Center, a conservative think-tank.

While seated together under the bright lights of the photographers, they spoke of their mutual interest in the peaceful uses of nuclear power, and a shared vision of locating such reactors underground for added safety. Sakharov then steered the discussion to arms control, reiterating his belief that deployment of a space-based ballistic missile defence would have a destabilizing effect on the world balance of power. For his part, Teller, who first suggested a Strategic Defense Initiative

Oncomouse released

Washington

Du Pont announced last week that it is to begin selling transgenic mice carrying activated human cancer genes, or oncogenes. The mice were developed by Philip Leder of Harvard University and Timothy Stewart, now at Genentech, and were the first genetically altered animals to be patented. The patent, held by Harvard University, covers any transgenic non-human animal bearing an activated oncogene sequence introduced by genetic-engineering techniques.

The first 'oncomice' will carry the *ras* oncogene, which has been shown to be common in a variety of human cancers, plus a mouse mammary tumour virus promoter which ensures that the oncogene is activated in breast tissue so that the mice develop a human breast cancer within a few months of birth.

Du Pont says that oncomice will help speed the search for new drugs to treat cancer by allowing laboratories to test drugs against a human cancer in a animal. Pharmaceutical companies are likely to be the major purchasers. A price of \$50–\$100 per mouse seems likely, five to ten times that of an ordinary laboratory mouse. Later next year, Du Pont will offer mice bearing *myc* and *neu* oncogenes which are also found in human cancers.

But the sale of Du Pont's mice may be threatened by changes in government policy. Patent applications for more than twenty genetically engineered animals have been filed with the US Patent Office but animal rights activists are mounting an increasingly vociferous campaign to ban the patenting of genetically engineered animals.

Alun Anderson

Boston nets BASF

Munich

THE chemical and pharmaceutical giant BASF AG announced on 11 November that a genetic engineering laboratory and pilot plant originally planned for West Germany will now be built in Boston. New regulations took effect in West Germany on 1 September, under which any production facility using genetically engineered organisms must be opened to the public for inspection before it can be approved. BASF anticipated objections from environmentalists and did not relish the idea of showing its design to the competition. Furthermore, the "climate for innovation" is better in Boston, said a spokesman, where there is a high concentration of experts in the field.

The complex is expected to cost DM100 million (about \$60 million) and to be completed in 1991. It will employ 60 scientists and 170 others in research and development in oncology and immunology. It was meant to be built at BASF headquarters in Ludwigshafen, near Heidelberg in southwest Germany.

Steven Dickman

Prospects for agreement on genetic resource issues

Washington

A BLUEPRINT for international action to preserve plant genetic resources emerged this week with the circulation of a draft document from a remarkable four-day international meeting held in Keystone, Colorado, in the summer. Although preservation of plant genetic resources is recognized as vital to the long-term health of the world's agriculture, attempts to co-ordinate efforts have been thwarted by competing international bureaucracies, and by divisive arguments between developed and developing countries.

The Keystone document represents a consensus that could cut a path through social and political complexities. The meeting organized by the Keystone Center in Colorado drew participants from all major concerned organizations, but encouraged them to speak as individuals. Only results that all could agree on were published.

The root of the problem is the fact that most of the world's naturally occurring genetic diversity for major crops exists in the developing world, but resources for breeding new crop lines or storing crop germplasm are in the developed world.

Proponents of 'farmers' rights' argue that plant breeders who use germplasm taken from primitive cultivated varieties ought to recompense the farmers who nurtured those seed lines for centuries. Developing countries doubt that genetic resources stored in banks in developed countries would be freely available.

Divisions between North and South also make it hard to decide who should be steering the international efforts to preserve plant genetic resources. In 1971, the World Bank established the Consultative Group on International Agriculture Research (CGIAR) which now acts as an umbrella organization for 13 international bodies, including the International Rice Research Organization in the Philippines and the Centro Internacional de Mejoramiento de Maíz y Trigo in Mexico. In 1974, another CGIAR body was established, the International Board on Plant Genetic Resources (IBPGR), with the aim of promoting worldwide conservation and utilization of plant genetic resources. IBPGR, like CGIAR, is run largely by the developed countries.

The Food and Agricultural Organization (FAO) of the United Nations is another major body interested in preserving genetic resources. In 1983, FAO established a new special Commission of Plant Genetic Resources. But several developed countries, notably the United States and Canada, declined to participate.

Several of the misconceptions barring

progress may have been cleared away by the Keystone meeting. A key accomplishment, according to Pat Mooney of the Rural Advancement Fund International, an outspoken proponent of farmers' rights, is the new understanding on compensation. Although compensation is owed to farmers, Mooney says, he now believes it is understood that it is not necessary to pay each individual farmer whose seeds are used in the development of a new line.

To resolve the compensation issue, participants in the Keystone dialogue suggest an international fund to be administered by a global independent advisory committee, with members drawn from FAO, IBPGR, the International Union for the Conservation of Nature and Natural Resources (IUCN), non-governmental organizations and industry.

There was also progress on who should guide international resource management. Although the FAO commission was praised for the work it had accomplished, the Commission was urged to expand its membership to include all countries. Barring that, the dialogue paper suggested forming a joint United Nations-FAO body that could report directly to both UN headquarters and FAO.

The report also urged that regional groups should become more involved in protecting resources, and that appropriate technologies be developed to make participation possible for the developing world.

Enthusiasm for the accomplishments of the Keystone dialogue is running high. M. S. Swaminathan, president of IUCN and chairman of the steering committee that organized the dialogue, says that considering the divergent opinions represented by the participants, the meeting went extremely well. Jose Esquinas-Alcázar, secretary of the FAO commission says that the Keystone meeting was the first real opportunity for all the interested parties to work together.

Some have reservations about just how much was accomplished, however. John Holden, a member of the IBPGR board of trustees, says the group's goals were so broad that it may be difficult to implement them. Donald Duvick of Pioneer Hi-Bred International, says that clearing the air is important, but difficult problems of intellectual property rights remain.

The dialogue does at least look set to continue. Swaminathan says there are tentative plans for another meeting in Leningrad next year.

Joseph Palca

The final report of the Keystone International Dialogue Series on Plant Genetic Resources (August 15–18, 1988), is available from PO Box 606, Keystone, Colorado 80435, USA.

Greater awareness of security in aftermath of computer worm

Boston & Washington

LIKE residents shellshocked after a spate of break-ins in their neighbourhood, computer network users at universities and research institutions around the United States agree that they have been forced into greater awareness of security issues by the recent attack of an unauthorized computer program upon thousands of computer systems around the country using the ARPANET, Milnet and NSF Net computer networks (see *Nature* 336, 97; 1988). The debate has now begun over whether increased security is needed in computer networks, and, if that is the case, how that will affect the free exchange of information.

If nothing else, the computer virus — or worm — that infected computers using a version of the UNIX operating system drove home how much computer networks have become part of the social structure of conducting science. Although the damage caused by the worm was slight, requiring no more than a fairly simple patch to electronic mail and related software source codes, most systems were removed from the networks for days while the repairs were made.

On electronic bulletin boards, the debate raged over whether Robert T. Morris, the Cornell University graduate student alleged to have come up with the invader was a 'hacker hero' or a criminal. Some believe that Morris uncovered significant flaws in the UNIX mail system, although others claim the flaws were well-known, and it was irresponsible to exploit them.

Users of the network seem to realize that it is not possible to construct a bug-proof system. Most suggestions for coping with future invasions involved raising the ethical standards of users, and simple but effective counter-measures such as requiring users to change their passwords frequently, and to use nonsense or random characters as passwords so that they cannot be easily deduced by clever programs.

James D. Bruce, professor of electrical engineering at Massachusetts Institute of Technology (MIT) and vice-president for Information Systems, says that several analysts and observers have begun to discuss tightening access to computer networks and even the possibility of closing down some of the networks. "Many people are hailing this as the end of the user-friendly era of computing, but if we've gone that far we have literally put our heads in the sand, or locked ourselves into a fortress we can't get out of", he says.

Last week in Washington, computer security experts from the Department of Defense and the National Security

Agency as well as other government agencies met to discuss what changes in network policy might be needed on a national level, but they have not yet released the results of their deliberations.

Bruce also says that teams at MIT have begun to look for programs that can be deleted from MIT networks. The ARPANET computer worm used a debugging program left on several versions of the 'sendmail' program on the Unix operating system to enter new computer systems. MIT's Project Athena computer network had already deleted this program from their version of "sendmail" and thus was largely spared from the virus attack. Bruce also stresses that his network managers are considering limiting access to certain programs to those with appropriate priority.

Despite these efforts, many researchers agree that the risks of attack through the networks are still significant. According to



Pascal Chesnais, research specialist at the MIT Media Laboratory and the user first to discover the worm at MIT, "It used to be that an isolated group was using the networks. They didn't need any passwords; there were no abuses. But things have changed." Chesnais likens the situation to people's homes, saying that people put locks on their front doors "mainly as a deterrent". If that proves insufficient, he says, "people begin to install alarm systems". Eventually, he says, it can get to the point where it is difficult to get into one's own home. Then, "you've clearly gone too far. Unfortunately, I'm afraid computer networks may get to be that way."

The Federal Bureau of Investigation (FBI) is investigating the worm attack to see if there were any violations of the Computer Fraud and Abuse Act of 1986. Last week, the FBI obtained search warrants for computer tapes with Cornell University containing "files from the Morris account including backups".

Seth Shulman & Joseph Palca

Mathematics and molecular biology

Berkeley

THE time is ripe for mathematics to make an impact on molecular biology, say the organizers of a University of California (UC) programme aimed at facilitating interaction between the two fields. The Program in Mathematics and Molecular Biology, based at the University of California at Berkeley, has received a \$2-million, five-year grant from the National Science Foundation to support collaborative research projects, fellowships and research conferences and workshops.

Although the programme is based at Berkeley, nine of the first eleven research projects it will support are scattered at different institutions around the United States. The projects focus on areas including DNA geometry and topology, protein-folding and genome mapping.

Sylvia Spengler, a biophysicist at the Lawrence Berkeley Laboratory and academic administrator for the programme, says that directions for future research will be evaluated with the aid of a five-member advisory board, which includes Genentech's vice-president for science, David Botstein, and biophysicist Aaron Klug of Britain's Medical Research Council.

The idea for the endeavour grew out of a collaborative study involving mathematical analysis of the geometry of DNA by Spengler, her biology colleague Nicholas Cozzarelli at Berkeley, and mathematicians Vaughan Jones at Berkeley, James White of the University of California, Los Angeles, and DeWitt Sumners at Florida State University.

Spengler and Cozzarelli realized that many areas of molecular biology could benefit from mathematics, but that collaborations might be slow to start because mathematicians and biologists are unfamiliar with work in each others' fields.

The programme was thus designed to act as a 'broker', said Spengler, and bring mathematicians and biologists together. At its Berkeley home base, it will keep several offices for visiting scholars who are likely to benefit from interactions with the Berkeley Center for Pure and Applied Mathematics and the Mathematical Sciences Research Institute.

Although the programme will be supporting relatively few researchers at any one time, Spengler said its research conferences and workshops will catalyse interactions among a broader circle of researchers. The first conference, planned for 1989 in Florida, will deal with applications of topology and geometry to molecular biology. Spengler hopes eventually to extend the programme's influence to European scientists by sponsoring joint workshops in conjunction with the NATO Research Series.

Marcia Barinaga

Mt Graham Observatory

SIR—I should like to clarify the status of the Mt Graham Observatory proposal (*Nature* 334, 188 & 645 and 335, 755; 1988). The site has been proposed by the University of Arizona (UA) as the location for future astronomical instrumentation. Legislation authorizing the establishment of a scientific research area, including astronomical facilities, was passed by the United States Congress on 20 October 1988.

The observatory proposal has been the subject of two environmental studies, neither of which identified any major effect except possibly on the endangered Mt Graham red squirrel. Two subsequent analyses by the US Forest Service, using recent survey data and computer models, have shown no significant effect on this possible subspecies; according to the Forest Service, about three squirrel locations might be lost. Most recently, the US Fish and Wildlife Service has proposed two options, both of which could accommodate the proposed observatory without jeopardizing the continued existence of the red squirrel. Both options permit immediate construction of the 10-m Submillimetre Telescope, the Vatican Advanced Technology Telescope and the 11.3-m Columbus Project and allow for subsequent expansion to seven instruments. Each option would affect two squirrels. In short, the proposed observatory would have a negligible effect on both the squirrel and the general environment. The University of Arizona is therefore confident that the observatory project can and should proceed on what appears to be the best available astronomical site in the continental United States.

More perplexing, amid all the concern for the red squirrel, is the lack of any serious comment on what is threatening its existence. The facts are, however, very clear. In the early 1940s, the non-native Abert's squirrel was introduced to this sky island site primarily for hunting purposes. It has now taken over all but 2,000 acres of the approximately 30,000 acres of former red squirrel habitat and outnumbers the native squirrel by roughly 20:1. It is clear that only one species can survive in the limited habitat provided by Mt Graham. Unless measures are taken, the red squirrel is probably doomed with or without the observatory. As far as we are aware, the UA has made the only specific recovery proposals so far offered. These were contained in its Habitat Conservation Plan proposed under Section 10 of the Endangered Species Act, which permits a proponent to develop and carry out conservation measures in conjunction with a project. These proposals have not been accepted.

In regard to the legislation creating a

scientific research area (150 acres) on Mt Graham, the proposed measure is consistent with US Fish and Wildlife Service and Forest Service recommendations. Indeed, according to the media and other reports, the agencies assisted in preparing draft legislation. Many precedents exist for such action by the US Congress, among them the establishment of the Lowell Observatory in Arizona and Langmuir Laboratories in New Mexico.

The UA is convinced that the proposed Mt Graham scientific research area is necessary for both astronomical and ecological research. It has also convincingly shown that the programme can be carried out without any significant deleterious effects on the environment or any major conflict with other users of the mountain. Indeed, Mt Graham, with its paved access road and public recreational activities, offers excellent opportunities for responsible and productive use.

PETER A. STRITTMATTER
(Director, Steward Observatory)

University of Arizona,
Tucson, Arizona 85721, USA

Reductionism

SIR—John Maddox¹ is quite justified in asserting that the accumulation of data may hinder the successful reduction by molecular biology of biological phenomena, a valid point which Bradbury² has totally neglected to comment upon.

Let us consider the field of mendelian genetics to serve as an example. For years, many scholars have claimed that molecular genetics is the reducing theory of mendelian genetics. Although superficially this seems to be so, the claim is far more complicated than it appears. Before a theory can be said to perform a successful reduction of another, the two criteria of the logical empiricist must be met. First, the general statements of the reduced theory (for our example, mendelian genetics) must follow by deductive logic from the laws of the reducing theory (for our example, molecular genetics). Second, the terms of the reduced theory must be either defined by, or consistently connected, to the terms of the reducing theory. Mendelian genes are defined in terms of their corresponding phenotypes. Mendelian genes therefore have three functions: mutation, expression and recombination³. During the 1970s and 1980s, however, molecular geneticists have extensively (but not exhaustively) researched each of these areas of gene function. Naturally, they have shown conclusively that there are very complex pathways involved in gene expression and recombination. Indeed, a mendelian gene cannot be fully identified with a 'mole-

cular gene'. Terms such as intron, exon, promoter, enhancer and suppressor all make impossible a one-to-one correspondence between the two 'genes', and therefore between the two theories. Thus, a many-to-many relationship seems to exist between mendelian and molecular genetics.

Although reduction is still logically possible, it becomes much more difficult to perform. Reductionists are obliged to come to terms with these many-to-many relationships. The most valiant attempt is that by Kenneth Schaffner⁴. He has corrected mendelian genetics and attempted to reduce the modified version through molecular genetics. Unfortunately, I believe that Schaffner's attempt results in an oversimplification of certain correspondences between the theories.

In short, reduction can occur only if the many-to-many relationship problem is resolved. However, as Maddox correctly points out, this particular task is becoming increasingly difficult as more terms are added to the molecular geneticist's vocabulary. The relationships between the two theories are becoming more complex; therefore, it is becoming more difficult to establish consistent relationships between different terms of each of the theories.

MYLES W. JACKSON
Department of the History and Philosophy
of Science and Wolfson College,
University of Cambridge,
Cambridge CB3 9BB, UK

1. *Nature* 333, 11 (1988).

2. *Nature* 333, 794 (1988).

3. Rosenberg, A. *The Structure of Biological Sciences* 95 (Cambridge University Press, 1985).

4. Schaffner, K. in *Conceptual Issues in Evolutionary Biology* (ed. Sober, E.) (MIT Press, Cambridge, 1984).

Science fellowships

SIR—Regarding the lower application rate for the fellowships offered by Japan's Science and Technology Agency (*Nature* 335, 287; 1988), I find it disturbing that Japan's bureaucracy is obsessed with pleasing the politicians of the capitalist Western nations. The stipulation that these fellowships are offered only to the nationals of the United States and ten industrialized countries in Europe and Oceania smacks of parochialism.

If applications for fellowships are considered internationally, disregarding the East-West, North-South labels, I am sure that there will not be any problem in filling the vacancies. Why should not Japan allow a budding Marie Curie from Poland or a cerebral Chen Ning Yang from China to enter its government laboratories?

SACHI SRI KANTHA
Laboratory of Marine Biochemistry,
Faculty of Agriculture,
University of Tokyo,
Bunkyo-ku, Tokyo 113,
Japan

New physics from new accelerators

Even before the foundations of the Superconducting Super Collider (SSC) have been laid in Texas, people are brooding about the next generation of particle accelerators.

IN the demonology of the particles of matter that should exist but which, as yet, do not, the Higgs boson (called H_0) is among the most conspicuous. Its predicated existence is a cornerstone of the grand unified theories of matter — a consequence of the asymmetries that mark the low-energy world we live in. Naturally, every new accelerator coming into operation raises people's hopes that, this time, that elusive particle will be found. And each new accelerator disappoints.

But is it possible that the accelerator people have been looking in the wrong places? Or in the right places in the wrong ways? That may be the implication of an argument now put forward by Richard Blankenbecler and Sidney D. Drell of the Stanford Linear Accelerator Center (*Phys. Rev. Lett.* **61**, 2324; 1988). To be fair, Blankenbecler and Drell are not so much concerned to cheer on the hunt for the Higgs boson as to guide the design of the generation of particle accelerators next after those now being built (LEP at Geneva and its equivalent at Serpukhov) or about to be (SSC in Texas).

Beyond those machines, all circular accelerators, there are likely to follow machines consisting of a pair of linear electron accelerators firing beams of electrons and positrons (or just electrons) at each other. To be worthwhile, the particles in each beam will have to be given energy greater than 500 GeV, ideally 1 TeV or more, which would not be possible in a circular machine because most of the energy put into acceleration would be promptly radiated away again as synchrotron radiation. Soviet physicists at Novosibirsk have already designed a linear collider of this kind — and have been frustrated by the decision that, when the time comes, it will be built at Serpukhov. In the United States and Europe, there are also people working away.

Blankenbecler and Drell are first concerned with a mechanism by which colliding beams of electrons and positrons are robbed of energy in more familiar accelerators, and which rejoices in the name (deliberately a near-pun) of beamstrahlung. In circular accelerators in which pulses of positrons and electrons are circulating in opposite directions, for example, pairs of particles will produce photons when they come close to each other even if their interaction is not so intimate as to produce other kinds of particles. As a consequence, it is customary to reduce the

loss of energy in those parts of a trajectory in which interactions are supposed not to occur by shaping the counter-circulating beams appropriately — giving each of them a ribbon-like cross-section, for example, so that their interpenetration is a minimum.

But what happens when the oppositely travelling beams are shaped not so as to minimize the loss of energy by this mechanism, but to maximize it? Then, the argument goes, there will be copious production of single photons, but the chance of producing pairs of photons simultaneously will be disproportionately enhanced. Is it possible that the process will even lead to the production of pairs of photons sufficiently energetic for their interaction to exhibit physical phenomena not otherwise easily observable?

The answer, it will have been guessed, is yes. The first step, the calculation of the chance that a single electron in a travelling pulse of, say, electrons will interact with a positron held, for the sake of the argument, fixed in space is a calculation for the text-books of quantum electrodynamics. What emerges from it is a function — essentially a probability distribution — specifying the chance that the moving electron will surrender a certain fraction of its momentum to radiation in the form of single photons. Strangely, calculating the chance that pairs of photons will be produced is then a classical probability calculation, partly because the repulsive Coulomb forces within a travelling pulse of charged particles repeatedly destroy the coherence of the motion. In the end, the calculation has to be numerical, but the results are clear-cut.

Briefly, the interaction of ribbon-shaped pulses will yield few cases in which pairs of energetic photons may interact with each other, but pulses with squat elliptical cross-sections will be relatively prolific sources of them. The essence of the argument is the conclusion that there should be circumstances in which careful shaping of the cross-section of the separate pulses of particles travelling in an accelerator will maximize the chance of photon-photon interactions by several orders of magnitude, four or five perhaps.

It is worth reflecting on the magnitude of what is proposed, which can be told from the numerical examples with which Blankenbecler and Drell illustrate their argument. As if in passing, they refer to electron pulses (travelling with a velocity

indistinguishable from that of light) such that, seen end-on, there are 10^{30} electrons or positrons per square centimetre of cross-section. They do not mention that, for practical purposes, the lateral dimension of the pulses will be measured in μm , not centimetres, but that is standard practice in electron accelerator laboratories. And the length (at rest in the laboratory frame) of each pulse will be rather less than a millimetre.

What, in these circumstances, will the copious production of energetic photon pairs produce? This is where the Higgs boson comes in. For, as a particle whose existence stems from the asymmetry in the real world between strong, weak and electromagnetic forces, H_0 must interact with almost everything in sight. In particular, even though it may lack electric charge, it should be produced as a single and recognizable entity in the interaction of two sufficiently energetic photons.

Part of the trouble with H_0 is that nobody knows how massive it may be relative to other particles, but there is at least a chance that, if the mass is less than those of the intermediate vector bosons mediating the weak nuclear interaction, the $W^\pm$ and Z^0 so far recognized only at CERN, examples may already have been produced, but have not been recognized in the welter of particles generated by other means.

But if the mass of the Higgs boson is greater than the equivalent of 100 GeV, there will be a greater chance of finding Higgs bosons in an electron collider whose beams have been appropriately shaped to enhance double-photon interaction than in the direct collision of oppositely charged electrons. Blankenbecler and Drell say that the obvious experiment, based on direct electron interaction, would be expected to yield 10 events a day, but that optimizing the beam shape to enhance photon collision should yield ten times as many.

The interest of this argument is not, of course, that it shows in what range of energy people should be looking for H_0 , or even that double-photon interactions are the mechanism by which it should be sought, but that the conventional wisdom of machine designers that beamstrahlung are an unmitigated nuisance may be mistaken. It will be interesting to see how quickly the designers respond by arranging to tune the shapes of the cross-sections of their pulses of particles.

John Maddox

Gene targeting

Getting nearer the mark

Brigid Hogan and Karen Lyons

HIGH on the wish-list of anyone studying mammalian cells is the ability efficiently to mutate any cloned gene whose function *in vivo* is not yet known. The general strategy for achieving this goal is to target a mutation to a given gene by homologous recombination between the endogenous gene and a mutant copy introduced into mouse embryonic stem (ES) cells in culture. When these cells are injected into a host blastocyst, they become incorporated into the normal embryo and can give rise to eggs or sperm that can transmit the mutation to the next generation (see figure). One of the main problems in applying this technique other than to model systems is to select the few ES cells in which homologous recombination has taken place from those in which the foreign DNA has integrated randomly into the genome. On page 348 of this issue¹, Mansour, Thomas and Capecchi help to bring realization of the dream a step nearer by describing an ingenious and general method for greatly increasing the efficiency of selection of correctly targeted ES cells.

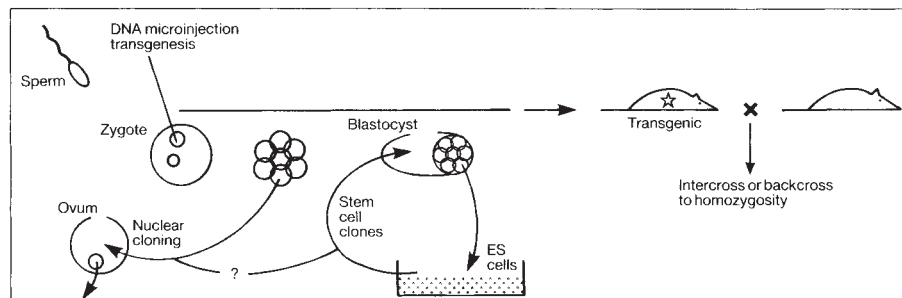
One of the targets chosen by Mansour *et al.* is the mouse *int-2* gene. Activated in certain virus-induced mammary tumours, *int-2* encodes a protein related to basic fibroblast growth factor, and is normally expressed only in a few specific regions of the mouse embryo^{2,6}. It is also transcribed at a very low level in undifferentiated ES cells, which was encouraging for the targeting experiments because it is believed that a gene may have to be in some kind of 'open' configuration for recombination to occur.

Mansour *et al.* used electroporation to introduce into cultured ES cells cloned DNA identical in sequence to 10 kilobases of the *int-2* gene. They incorporated two other pieces of DNA into this fragment. One, the bacterial gene encoding neomycin resistance (*neo'*) driven by a promoter known to be active in ES cells, both mutates the *int-2* gene (as it is placed so that it disrupts the coding sequence of one of the exons) and allows positive selection. The other, serving as a negatively selectable marker, is the herpesvirus-thymidine kinase (*HSV-tk*) gene, which is positioned at the end of the linear DNA fragment. Initially, cells in which the foreign DNA has integrated anywhere in the ES genome are selected by growth in the presence of the drug G418; only those cells with a *neo'* gene will survive. These cells are then treated with the nucleoside analogue gancyclovir (GANC), originally developed as an antiviral agent because it is converted to a toxic product by

HSV-tk and not by the cellular enzyme. If homologous recombination has taken place, the *HSV-tk* gene will have been excluded from the ES genome because it lies outside the region of homology between the vector and the target locus. If, on the other hand, integration has been random, then the chances are high that an intact viral *tk* gene will be present in the genome and the cells will be killed by GANC. Of the 1.6×10^5 G418-resistant colonies they obtained in the first round of selection, Mansour *et al.* find that 81 are also GANC resistant, a 2,000-fold enrichment. They then tested the 81 colonies for homologous recombination by Southern analysis of their genomic DNA. Four of the 81 give the pattern of restriction fragments predicted if the targeting vector (minus

with the incoming vector. *Taq* DNA polymerase is then used to amplify the sequence between the two probes, generating a DNA fragment of predicted size only if homologous recombination has taken place. Pooled cells giving a positive signal then have to be rescreened several times until a clonal population is obtained. Despite the relative tediousness of this method, which does not have the benefits of direct selective enrichment, Joyner and Rossant have now produced chimaeric mice containing cells with a correctly targeted *en-2* gene. In these mice, the targeted cells intermingle with normal cells.

What will happen when, in subsequent generations, the whole embryo is made homozygous for the null mutation generated by *neo'* insertion? In the case of the *int-2* and *en-2* genes, the embryos may show quite specific defects in development because the genes are normally expressed in very restricted regions^{4,5,7}. But for other genes expressed by a wider range of cell types, it might be desirable to



A series of possible routes to transgenesis in mammals. DNA microinjection provides the straightforward route. The alternative route, using cultured embryonic stem cells, provides a means by which the full experimental genetic manipulation in culture, including targeted gene mutagenesis, may be used before reconstruction of the transgenic animal. The other possible route shown, that of nuclear cloning, may be a more useful method in larger mammalian species than mice. (Courtesy of Martin Evans.)

HSV-tk) had correctly inserted into the *int-2* gene.

One advantage of the GANC enrichment technique is that it requires little information about the DNA sequence of the targeted gene. This is in contrast to the polymerase chain reaction (PCR) technique being used in other laboratories for distinguishing between homologous and random integration. For example, Alexandra Joyner and Janet Rossant (Mount Sinai Hospital, Toronto) reported at a recent meeting* their strategy for targeting the mouse *en-2* homoeo-box gene. This involves introduction into ES cells of a cloned *en-2* fragment in which *neo'* disrupts the coding sequence near the 3' end. Genomic DNA is then isolated from pools of G418-resistant cells and hybridized with two short oligonucleotide probes, one located close to the end of the *neo'* gene and one within the endogenous gene, just downstream of the boundary

introduce more subtle mutations in which small domains of the protein are altered. For this purpose, the technique reported at the meeting by Andreas Zimmer and Peter Gruss (Max Plank Institute, Göttingen) may be more appropriate. They have microinjected into ES cells part of the mouse *Hox-1.1* gene in which a short oligonucleotide disrupts the coding sequence near the 3' end. Instead of introducing a stop codon, the oligonucleotide insert could equally well be designed to produce a specific amino-acid change. Because the starting population of microinjected cells is small, screening by PCR can be carried out relatively quickly without the need for prior selection, and the frequency of homologous recombination can be as high as 1 in 250 cells injected.

ES cells may lose their ability to colonize the germ line after prolonged culture or unfavourable conditions. The news is therefore eagerly awaited that the chimaeric mice generated in the different laboratories have successfully transmitted their

* *Mouse Molecular Genetics*, Cold Spring Harbor, New York, 24-28 August 1988.

targeted genes to the next generation. Among those particularly anxious for a positive result are those studying mammalian embryos, who are now in the fortunate position of having many cloned genes encoding structural motifs such as homoeo boxes, zinc-finger DNA-binding domains and growth-factor homologies also found in *Drosophila* gene products known to be involved in spatial organization and differentiation. Although the patterns of expression of the mouse genes are consistent with a similar role in mammalian development, definitive evidence can come only from analysis of mutants. Gene targeting in ES cells thus offers exciting

possibilities for mammalian developmental genetics. □

1. Mansour, S.L., Thomas, K.R. & Capecchi, M.R. *Nature* **336**, 348–352 (1988).
2. Nusse, R. *Trends Genet.* **4**, 291–295 (1988).
3. Dickson, C. & Peters, G. *Nature* **326**, 833 (1987).
4. Wilkinson, D.G., Peters, G., Dickson, G. & McMahon, A.P. *EMBO J.* **7**, 691–695 (1988).
5. Jakobovits, A., Shackleford, G.M., Varmus, H.E. & Martin, G.R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7806–7810 (1986).
6. Smith, R., Peters, G. & Dickson, C. *EMBO J.* **7**, 1013–1022 (1988).
7. Davis, C.A., Noble-Topham, S.E., Rossant, J. & Joyner, A.L. *Genes Dev.* **2**, 361–371 (1988).

Brigid Hogan and Karen Lyons are in the Department of Cell Biology, Vanderbilt University Medical School, Nashville, Tennessee 37232, USA.

Cosmology

Back to the future

John L. Friedman

"THIS must be a simply enormous wardrobe." Lucy came by chance across the boundary between our world and Narnia. Seen from either world, or so I imagined when I read C.S. Lewis's *The Lion, the Witch and the Wardrobe* (Macmillan, New York, 1950), the wardrobe's walls seem to enclose a small room, but a different universe extends out from each side. The wardrobe is thus a 'wormhole', a topological tunnel that joins two universes or two distant regions of a single space-time. Although wormholes are solutions to Einstein's equations, if they are built from ordinary matter they collapse too quickly to let travellers through and crush anyone foolhardy enough to try. At least over short distances, however, the behaviour of matter is not ordinary, and new results of Michael Morris, Kip Thorne and Ulvi Yurtsever (*Phys. Rev. Lett.* **61**, 1446–1449; 1988; also Morris, M.S. & Thorne, K.S. *Am. J. Phys.*; in the press) suggest that the laws of physics might permit an advanced civilization to keep a wormhole open.

But now we find ourselves on dangerous ground. If a civilization can maintain wormholes, a simple argument seems to show that it can convert them into time machines as well. That is, a wormhole can apparently be altered to connect one place in the universe to a place in its past. The argument is carefully crafted, but the conclusion is difficult to accept. Have we missed a fundamental constraint that maintains the consistency of physics and prevents us from journeying back in time, for example, to murder our ancestors?

A wormhole is easily pictured by imagining that space is two dimensional, as in Fig. 1. Like the wardrobe walls, a wall built along the circle C would seem from either side to enclose a small area, but a traveller crossing the wall would see a new universe open out in front of her. If constructed from space alone, or from

matter with positive energy, all wormholes are black holes, surrounded by gravitational horizons (the dotted circles) that act as one-way membranes. Although objects can fall into such a wormhole, they cannot get out. Once inside, the matter collapses on itself, and the wormhole pinches together.

Carl Sagan's novel *Contact* (Simon and Schuster, New York, 1986) describes a wormhole constructed long ago by an ancient, extinct civilization as part of a rapid galactic transit system. Hoping to keep his science fiction honest, Sagan sent a preliminary draft to Kip Thorne, who was, of course, aware of the theorems forbidding traversible wormholes. The theorems rely on the positive energy of ordinary matter, and Morris, Thorne and Yurtsever set out to see if one could avoid wormhole collapse by exploiting the fact that matter is described by quantum fields.

Although the total energy of a field is positive, on small scales it fluctuates rapidly in space and time, and almost all quantum states include regions with negative energy density. Even the quantum vacuum is a busy place, in which particles quickly appear and then vanish. Its zero average energy density simply counts an equal chance for each small region to have positive or negative energy. Moving two metal plates close together changes the vacuum between the plates by preventing the electromagnetic field from creating photons with wavelengths longer than plate spacing. The lack of these photons

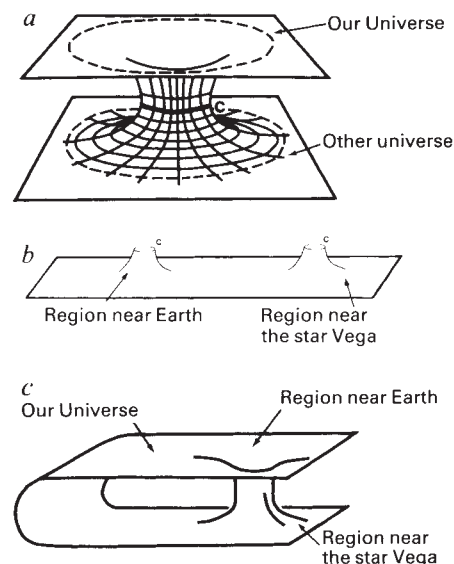


Fig. 1 a, Two 2-dimensional universes joined by a wormhole; b and c, distant parts of a single universe (say Earth and Vega in our Universe) similarly joined. The two circles (c) in b represent the same set of points, and by sewing them together one obtains c.

lowers the energy between the plates — and because it started at zero, it ends up negative. By surrounding each end of a wormhole with a spherical metal plate, one produces enough negative energy between the plates to avoid a horizon and to prevent the collapse of the wormhole.

The authors claim only that the calculation is intriguing, not that it is correct: the detailed structure of the plates themselves is ignored and it could conceivably be crucial. At the least, they show that there is no obvious reason why one cannot maintain a wormhole.

To convert the wormhole into a time machine does not seem to be too hard. Recall that in special relativity, two observers at rest relative to one another have clocks that run at the same rates, whereas an observer who accelerates relative to inertial (unaccelerated) clocks, turns around, and returns, will find after her round trip that the inertial clocks read a later time than her own.

Now consider clocks at each end A and B of a wormhole, each reading the same time. In its own part of the universe, each wormhole end behaves like a massive object. By placing a heavy asteroid near end B and accelerating it one can use the asteroid's gravity to drag the wormhole end along a round trip. The universe clocks, the inertial clocks outside the

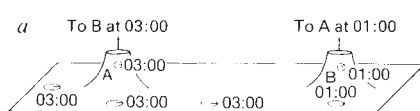


Fig. 2 Time machine — a wormhole joins the part of the universe near B at 03:00 universe time to the part of the universe near A at 01:00 universe time.

wormhole, are unaffected by the trip; they remained at rest relative to each other and to end A. Similarly, clocks at A and B inside the wormhole were at rest relative to each other while end B was moving relative to the universe clocks around B; thus, inside the wormhole at ends A and B are clocks that read the same time, say 01:00.

After its round trip, less time has elapsed on the clock at B than on the surrounding universe clocks, so the universe clocks near B read a later time than B, say 03:00. But the universe clocks surrounding A read the same time as clock A, namely 01:00. In other words, end A is attached to a part of the universe that is two hours earlier than the universe surrounding B (Fig. 2). If a traveller can go from A to B the long way in one hour, and through the handle in a few seconds, then by starting at A at 02:00 universe time, she will reach B at 03:00 universe time. But when the universe clocks near B read 03:00, B's clock reads 01:00. The traveller will enter the wormhole at 01:00 on B's clock and emerge a few seconds

later to see A's clocks and the surrounding universe clocks read 01:00, one hour earlier than she started. Overall, she will have travelled backwards in time from 02:00 to 01:00.

Strictly speaking, once there is any particle path that takes the particle along a closed loop to its starting place at the same time, one cannot claim to predict the future, because the future can change the past. So one cannot really predict, from the laws of physics, that a time machine can be constructed in this way (or in any way). Morris *et al.* designed their machine to get around this problem, by adjusting the particle paths that return to the past so that the particles carry too little energy to the past to disrupt the construction of the machine. I think they have got that part right. Then what prevents a traveller from journeying to the past to murder her ancestors is a question more of science than fiction. □

John L. Friedman is in the Department of Physics, University of Wisconsin, Milwaukee, Wisconsin 53201, USA.

Evolution

A simple measure of complexity

Rolf Landauer

"So God created man in his own image, in the image of God he created him; male and female he created them" (ref. 1). It is apparent that 'man' is in a way more complex than most of what surrounds him; but can that intuitive distinction be made analytically? Seth Lloyd and Heinz Pagels² have given us a new attempt to define complexity. It is one of the remarkably few thrusts in this area which is not conspicuously vacuous, and deserves serious consideration.

Physical scientists are becoming increasingly concerned with the closely related questions of complexity and of self-organization. This concern is often motivated by the attempt to gain insights into the origin of life and the unfolding of evolution, and to do so without getting dirty with the realities of biology. Much of this endeavour promises too much for too little work, and must be regarded with the same suspicion as an offer to sell the Brooklyn Bridge for a dollar. Consider, as an example, a book³ which has *Complex Systems* in its title and *Self-organization* in its subtitle. We find much in this book about the kinds of transition which occur when a Bénard instability of a fluid is excited, a laser turned on or a chemical oscillation excited. But a simply describable system, even if chaotic, going through its entertaining but ordained behaviour, is not self-organization, any more than the ticking of a pendulum clock or the rotation

of a water wheel driven by a stream.

The new measure proposed by Lloyd and Pagels is closely allied to earlier attempts by C. Bennett^{4,5} and by H. Kuhn⁶. All three formulations deal with the whole sequence of events that lead to the object whose complexity is described; the atomic structure or the blueprint specifying the object is not enough. These definitions are, in a sense, tautologies. They all roughly say: that which is reached only through a difficult path is complex. Tautologies, however, are welcome if they replace nonsense. Darwin cleared the air by telling us that the survivors survive. The emphasis on the path leading to an object, or state, has also been a central theme in my work⁷. If we want to know whether an ecology with two-tusked elephants is more or less likely than one with four-tusked elephants, we cannot do that by just comparing these two states, but must consider the paths leading to them. Simple thermodynamic or information-theoretic criteria, applied to the states to be compared, are inadequate.

Right or wrong?

A quantitative measure is best justified through utility in further application. None of the work in this area has gone that far. Thus, it becomes a matter of taste whether a proposed measure is judged to capture an intuitive notion. We cannot decide objectively whether the definition

is right or wrong. Indeed, there is some danger that the concern with the formulation of a definition comes at the expense of clearer questions. For example, does a system which can give rise to a complex evolution need a minimal degree of departure from thermal equilibrium or a minimal degree of spatial heterogeneity? As Bennett has asked (personal communication), if we consider a homogeneous fluid existing between two large plates 1 cm apart, with the upper plate 1 K hotter than the lower one, can we expect a civilization with heat engines eventually to develop in the layer?

Lloyd and Pagels define the 'depth' of a state reached by a particular trajectory to be $-k \ln p_i$, where p_i is the probability of the i^{th} trajectory. For the whole range of possible trajectories, the resulting weighted average is $-k \sum p_i \ln p_i$. The p_i represent probabilities which are consistent with all the measurements that have been made on the system during its history. The arbitrary multiplicative constant k is chosen to be Boltzmann's constant for systems whose successive configurations can be described in the physical space of statistical mechanics. The authors point out that in this case the depth becomes the entropy passed on, during the evolution, to degrees of freedom other than those needed in the specification of the final state. This quantity is shown to satisfy some plausible requirements: for example, the depth of the evolution resulting from two successive stages is equal to the sum of the separate depths. Lloyd and Pagels allow for the possibility that one may not want to count all the entropy passed on during the evolution, but only that connected with the interesting parts of the process. But in that case we are left with a very subjective definition.

Kuhn⁶, in a long sequence of papers which are particularly close to the new work², focuses more explicitly on the origin of life and evolution, and defines 'knowledge' gained as the number of bits discarded along the way. He measures the number of evolutionary choices that have been made, and is concerned with genetic bits, or more generally, with bits characterizing the description of the states along the way. In contrast to the principal thrust of Lloyd and Pagels, Kuhn does not include all the uninteresting entropy generated by an organism's metabolism.

Bennett⁴, instead, defines the depth of a computational target as the minimum number of steps required in the computation, starting from the most compact possible program. This is a definition which has some chance of generalization and application to a more general physical process, rather than just to a computation^{5,8}. The running time for a computation has, of course, been an object of great interest in computer science, where there

is a need to know the relative efficacy of algorithms. Bennett, instead, uses it as a measure of complexity of the computational target. Bennett's measure is complementary to algorithmic entropy⁹, which characterizes the length of the minimum program leading to a target, and is the basis of modern discussions of randomness.

Historical trail

Under the measure proposed by Lloyd and Pagels, two cats and a kitten are not very much more complex than two cats; reaching the state with two cats was the difficult part. The additional emphasis the authors place on the actual historical trail, rather than the most likely history (to take one alternative possibility) does not seem to be an unqualified asset. Stone fragments known to be the result of human intervention are burdened with the whole history of human evolution by this approach, and are assigned a much greater complexity than the same fragments would have as the result of a natural geological event.

Consider a sequence of steps leading to the origin of life in a shallow pool near the base of a waterfall. For the combination of pool and fall, the entropy generated by the powerful fall dwarfs that of the more interesting events in the pool. Is the waterfall really more complex than the molecular events in the pool?

This new work represents a collaboration between S. Lloyd and the late H. Pagels, who was killed this summer in an accident while they were mountain climbing together. Pagels's own trail of evolution toward the views expressed in ref. 2 is provided by a fossil-like record in the literature. A conference¹⁰ he organized in 1983 included a talk by Bennett on logical depth. In a later book review¹¹, he evaluates Prigogine's approach, saying "... all too often that which is correct is not new and that which is new is not correct." In one of his popularizations¹², Pagels tries to come to grips with complexity in a very broad way, but still without a detailed analytical approach. And finally in ref. 2, written with Lloyd who followed an independent trail leading to this question, the detailed formulation appears. □

1. Genesis 1:27

2. Lloyd, S. & Pagels, H. *Ann. Phys.* **188**, 186–213 (1988).

3. Serra, R., Andretta, M., Zanarini, G. & Compiani, M. *Introduction to the Physics of Complex Systems* (Pergamon, Oxford, 1986).

4. Bennett, C.H. *Found. Phys.* **16**, 585–592 (1986).

5. Bennett, C.H. in *Emerging Syntheses in Science* (ed. Pines, D.) 297–313 (Addison-Wesley, Massachusetts, 1987).

6. Kuhn, H. *IBM J. Res. Dev.* **32**, 37–46 (1988).

7. Landauer, R. in *Self-Organizing Systems* (ed. Yates, F.E.) 435–443 (Plenum, New York, 1987).

8. Wright, R. *The Sciences* **25**, (2), 10–12 (1985).

9. Chaitin, G.J. *Scient. Am.* **259**, 80–85 (July 1988).

10. Pagels, H. (ed.) *Computer Culture* (New York Academy of Science, New York, 1984).

11. Pagels, H.R. *Phys. Today* **38**, 97–99 (1985).

12. Pagels, H.R. *The Dreams of Reason* (Simon & Schuster, New York, 1988).

Rolf Landauer is in the IBM Research Division, T.J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA.

Archaeology

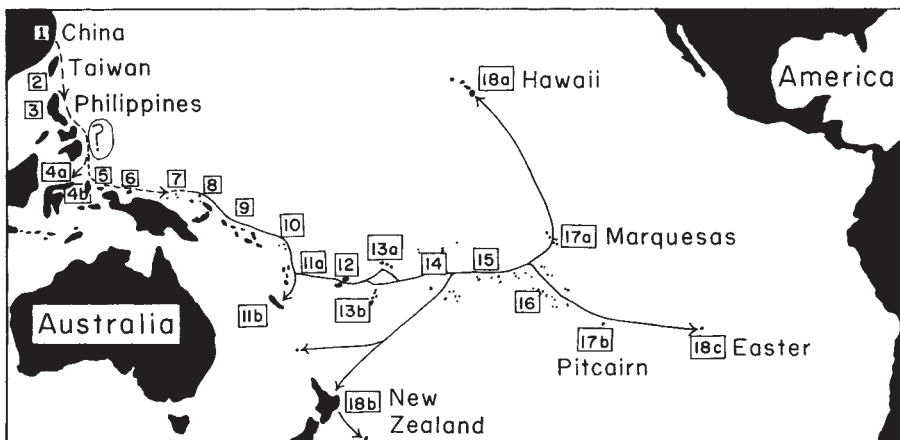
Express train to Polynesia

Jared M. Diamond

POLYNESIA was the last area of the world to be settled by neolithic humans, who had to overcome unprecedentedly wide water gaps in colonizing the islands. Identifying the dates and routes of that colonization should help elucidate why modern Polynesian languages differ, how an ancestral society radiated into varied daughter societies on different islands, and how neolithic people navigated between the islands. Recent excavations by Patrick Kirch and others on the Mussau islands of the Bismarck archipelago in the south-west Pacific Ocean illuminate the origin of Polynesians (Kirch, P.V. & Hunt, R.L.

the 'express train to Polynesia' theory assumes instead that Lapita culture arrived full-blown in the Bismarcks, having arisen far to the west in island south-east Asia and spread rapidly eastwards. The economic basis of Lapita is also debated: were Lapita peoples initially 'beach-combers' living off marine resources, or did they already have intensive agriculture and the main domestic animals (pig, dog and chicken) of historical Polynesians? Kirch's excavations on the Mussau islands may help resolve both of these debates.

The excavations uncovered the largest-known Lapita site. The many potsherds



The express train to Polynesia. Stations 8–13 are attested by Lapita pottery, Talasea obsidian, or both; stations 9–17a and 18, by extant Polynesian populations; 17b, by Polynesian artefacts; and 3 and 4a, by pottery resembling Lapita. 1, 2, and 4b–7 are speculative. 4a, Sulawesi; 4b, Halmahera; 5, Waigeo; 6, Biak; 7, Wuvulu; 8, Bismarcks; 9, Solomons; 10, Santa Cruz; 11a, Vanuatu; 11b, New Caledonia; 12, Fiji; 13a, Samoa; 13b, Tonga; 14, Cooks; 15, Tahiti; 16, Tuamotus.

Radiocarbon **30**, 161–169; 1988; Kirch, P.V. *Natn. geog. Res.* **4**, 328–342; 1988; Kirch, P.V. *J. Field Arch.* **14**, 163–180; 1987).

Polynesian culture developed in the island groups around Samoa and Tonga from a precursor culture termed Lapita and known for its characteristic pottery and other artefacts. Lapita culture flourished from about 1600 to 500 BC on Pacific islands from the Bismarcks east to Samoa. Lapita people were certainly the first humans to reach Fiji, Samoa and Tonga, and possibly New Caledonia and the New Hebrides as well.

Despite the wealth of recent archaeological discoveries relating to the Lapita, the origins of this culture are still unknown (see figure). One theory (see, for example, Anson, D. *Archaeol. Ocean.* **21**, 157–165, 1986) postulates that Lapita evolved gradually in the Bismarcks themselves, where humans are known to have arrived from the west at least 33,000 years ago. A competing theory (for example, Bellwood, P. *Prehistory of the Indo-Malay Archipelago*, Academic, Sydney, 1985),

found there date back to 1600 BC — essentially the same dates as the earliest Lapita culture in the New Hebrides, New Caledonia, Fiji, Samoa and Tonga. The artefacts found at the site are full-blown Lapita: there is no trace of gradual local origins, and the oldest deposits contain the most elaborate Lapita pottery. Thus, these finds support the 'express-train' model, in which Lapita arrived from the west and spread rapidly eastwards without measurable pause. The Lapita must have covered 4,500 km of unexplored ocean and islands separating the Bismarcks from Samoa in a mere few centuries or less.

As a remarkable bonus, the Mussau sites included an old shoreline where a brackish aquifer had preserved organic remains anaerobically. Among the preserved materials were wooden post bases for a stilt house over a lagoon, this documenting ancient origins for a house style now widespread from island south-east Asia throughout Melanesia. The materials also include fragments of coconuts and edible nuts or fruits of at least 18 other tree

species now important in oceanic agriculture. Root crops like taro and yams would not have been preserved as they lack woody tissue, but finds of shell peelers and scrapers suggest that these oceanic root crops were also cultivated. In addition to bones or shells of fish, turtles and molluscs, the site yielded bones of the three main Polynesian domesticated animals. Thus, the Lapita arrived in the Bismarcks with agriculture and animal husbandry.

If the passengers on the 'express train' were not only navigators but also farmers and fishermen, three questions become critical. First, no evidence of pre-Lapita settlement was found on Mussau, yet New Guinea and the large central Bismarck islands of New Ireland and New Britain have been settled for at least 40,000, 33,000 and 11,000 years, respectively (Allen, J. *et al.* *Nature* **331**, 707-709; 1988). These large, high islands are visible from each other and separated by narrow water gaps, whereas Bismarck outliers like Mussau and Manus are more isolated and invisible from other land. Did Pleistocene colonists reach only the largest visible targets, leaving Bismarck outliers to be discovered tens of thousands of years later by Lapita explorers?

Second, what happened after the Lapita wave swept out to Samoa around 1600 BC? Not until nearly 2,000 years later is there any securely dated evidence of human presence in Polynesia beyond Samoa. In one view (Jennings, J. *The Prehistory of Polynesia*, Harvard University Press, Cambridge, 1979), this 'long pause' in Polynesian dispersal is real and represents time required for discovering islands beyond the much wider seaways east of Samoa. An alternative is that the long pause is at least partly an artefact of inadequate archaeological exploration east of Samoa (Kirch, P.V. *J. poly. Soc.* **95**, 9-40; 1986). Did the express train really grind to a halt in Samoa and, if so, why and for how long?

Finally, where west of the Bismarcks did the train start, and what were its intermediate stations? Pottery and other artefacts similar to Lapita objects have been found in the Philippines and Sulawesi. If Lapita culture arose there, there were probably at least four stops *en route* to the Bismarcks (see figure): the northern Moluccas (Halmahera and Morotai), Waigeo and neighbouring islands; Biak and other islands of Geelvink Bay; and the Wuvulu/Hermit/Ninigo groups. Thus, archaeological excavations in Indonesia and Papua New Guinea are now required to trace initial steps in a journey that carried neolithic seafarers 15,000 km across the Pacific, from island south-east Asia to Easter Island. □

Jared M. Diamond is a professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Neurotransmission

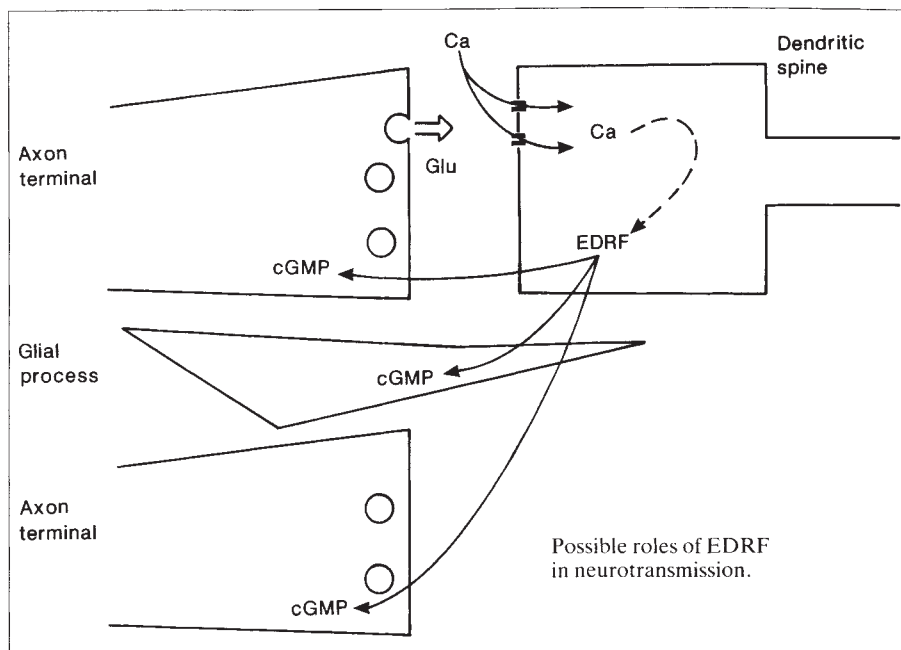
A relaxing factor in the brain

Charles F. Stevens

THIS is a strange story, but an interesting one. Garthwaite *et al.* show on page 385 of this issue¹ that the same unusual communication system that operates between endothelial and smooth muscle cells in blood vessels — a system whose discovery and acceptance has set vascular physiology on its ear² — also works in the brain. The authors of this new work give evidence that a factor, called endothelium-derived relaxing factor (EDRF) by vascular physiologists, is released through activation of glutamate receptors in the cerebellum and triggers an increase in the

other hand, inhibits EDRF.

Garthwaite *et al.* in their new work show¹ that the same thing happens in suspensions of cerebellar cells stimulated with the excitatory neurotransmitter glutamate. Glutamate causes the release of a factor that leads to an increase in cGMP concentrations. The authors identify the factor as EDRF because it is released into the medium suspending the cells isolated from cerebellum, is protected by SOD, is inhibited by haemoglobin, and will cause smooth muscle in aortic strips to relax. Furthermore, the effect of glutamate is



level of the second messenger cyclic GMP in the neighbouring cerebellar cells.

To understand what could be happening in the brain, we first need to review the role of EDRF in the vascular system². Acetylcholine, histamine and other factors cause the smooth muscle cells of blood vessels to relax. The natural assumption was that these agents act directly on the vascular smooth muscle, but it is now believed that the relaxing effect is mediated by endothelial cells: acetylcholine and similarly active agents cause endothelial cells to release the factor EDRF, and EDRF in turn diffuses to the smooth muscle cells where it activates guanylate cyclase to make the cGMP that produces relaxation. EDRF has been identified, not yet to everyone's satisfaction, as the very simple molecule nitric oxide (NO). Superoxide dismutase, an enzyme with the unfortunate acronym SOD, protects the very labile EDRF (NO?) and thereby enhances its effect. Haemoglobin, on the

mimicked by sodium nitroprusside, an agent that, like the clinically used vasodilator nitroglycerine, causes NO to be formed and acts like EDRF on vascular smooth muscle.

One of the most intriguing aspects of this story relates to the glutamate-receptor end of the process: glutamate-receptor channels mediate most if not all excitatory synaptic transmission in the brain and come in two main varieties, NMDA and non-NMDA types³. NMDA, or *N*-methyl-D-aspartate, is a glutamate analogue not itself found in brain, which selectively activates, and thereby defines, a particular class of glutamate receptors. NMDA receptors have attracted much attention lately because they have been implicated in long-term potentiation (thought to be involved in learning⁴), activity-dependent rewiring of neuronal connections during development^{5,6}, and excitotoxicity⁷, a pathophysiological process common to many disease states in which neurons are

killed by excessive exposure to glutamate. In all of these processes, an influx of calcium ions through the NMDA-receptor channel, and the consequent increase in calcium concentration in the spine process bearing the postsynaptic membrane, are thought to be key factors.

In addition to NMDA's involvement in these processes, Garthwaite *et al.* now show that NMDA receptors participate in the release of EDRF from cerebellar cells. These authors find that NMDA, like glutamate, causes EDRF to be released, that the response is blocked by the specific NMDA antagonist aminophosphonovaleate, and that calcium ions must be present in the extracellular medium for NMDA to produce EDRF release (just as calcium must be present for long-term potentiation and excitotoxicity to occur). The idea, then, as shown in the figure, is that glutamate activates NMDA-receptor

channels which permit an influx of calcium ions. These calcium ions in turn activate something, perhaps an enzyme that releases NO from the guanidino group of arginine, to cause EDRF production and release. Finally, the EDRF diffuses to the target cells where it causes the activation of guanylate cyclase and the increase in cGMP concentration. What the cGMP does remains a mystery. But laboratories that work on long-term potentiation are already looking to see if EDRF has any effects on this process. □

1. Garthwaite, J. *et al.* *Nature* **336**, 385–388 (1988).
2. Vanhoutte, P.M. *Nature* **327**, 459–460 (1987).
3. Stevens, C.F. *Nature* **322**, 210–211 (1986).
4. Kennedy, M.B. *Nature* **335**, 770–772 (1988).
5. Cline, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 4342 (1987).
6. Kleinschmidt, A. *et al.* *Science* **238**, 355–358 (1987).
7. Choi, D.W. *Trends neurosci.* **11**, 465–469 (1988).

Charles F. Stevens is in the Section of Molecular Neurobiology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA.

Quasar distribution

Not too close for comfort

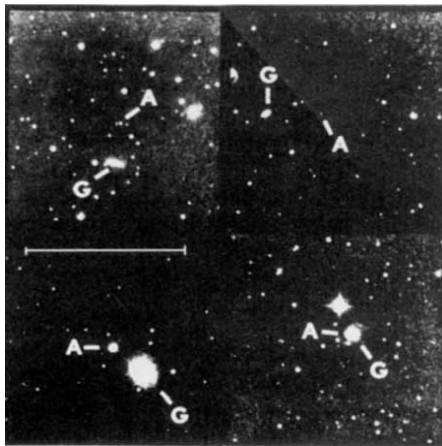
Claude R. Canizares

ARE quasars, commonly thought to lie at the farthest reaches of the Universe, preferentially aligned with galaxies near our own? This question, asked frequently over the past 25 years, is raised again by Webster *et al.* on page 358 of this issue¹. Their answer is yes. Similar claims have been made repeatedly by a few iconoclasts who have exhibited their own findings as proof that the quasars are really much closer than generally thought and, therefore, that standard cosmology is fatally flawed. In contrast, the new result is interpreted entirely within the standard theory; the excess background quasars aligned with foreground galaxies are yet another manifestation of gravity acting as a lens and give further evidence for large amounts of dark matter surrounding every visible galaxy².

To Pythagoras, it seemed that all the stars in the heavens lay on a single spherical surface. So it appears even to modern astronomers, who have no direct way of measuring the distance of anything but the closest stars. In standard cosmology, the distance of a galaxy or quasar is deduced indirectly, from its velocity as revealed by the Doppler shift (redshift) of its radiation. Its velocity locates the galaxy in the large-scale flow of the expanding Universe and fixes its distance in terms of a measurable quantity, the Hubble constant. The distances derived for many quasars are several tens of billions (10^{10}) of light years, far beyond any visible galaxies and near the edge of the observable Universe. Such large distances imply that quasars, which are no bigger than our Solar System, are hundreds of times more luminous

than a whole galaxy of stars.

An increasing body of evidence leads most astronomers to accept these extreme values of distance and luminosity. But a few observers have repeatedly questioned the wholesale extrapolation of the redshift–distance relation^{3,4}. They support their objections with examples of high-



Four examples of fields in which quasars (A) and foreground galaxies (G) are almost aligned. Scale bar, 5 arcminutes. (From ref. 5.)

redshift quasars that lie very close on the sky to an obviously nearby galaxy. The argument turns on statistics: are such alignments too numerous to be explained as chance superpositions of two uncorrelated samples, foreground galaxies and background quasars? If so, then at least some of the quasars must be physically associated with nearby galaxies. Equating association with proximity means that these quasars must be thousands of times closer to us than their redshifts imply in

the standard cosmology. Most astronomers remain unconvinced, finding these statistical arguments unsubstantiated and idiosyncratic.

Webster *et al.*¹ meet the statistical difficulties head-on and make a strong case for the preferential alignment of some nearby galaxies with presumably more distant quasars. They use an automated machine of marvellous precision that scans astronomical photographs and apply clever algorithms to pinpoint candidate quasars, whose identities they subsequently confirm. This allows them to amass a large sample based on objective criteria and presumably free of the suspected observer-selection defects of most earlier studies. Out of 296 quasars, they find 11 that are closely aligned with foreground galaxies. Such a large number, they compute, has less than 0.01 per cent probability of occurring by chance. Two other recent, analogous surveys, for quasars selected by their X-ray (see figure) or radio emission, also show excess alignments, although with lower statistical weight^{5,6}.

These results do not, however, undermine the redshift–distance relation. Gravitational lensing, the magnification of a background object by the gravitational field of an object in the foreground, provides a mechanism for emphasizing such galaxy–quasar alignments within the context of standard cosmology^{7–9}. The preferential alignment occurs when a single object in the foreground galaxy — star, giant planet or black hole — lies within about 10^{-6} arcseconds of our line of sight to a distant quasar (the few known cases of lensing by entire galaxies, discussed recently by J. A. Tyson², are related but different). The quasar is magnified, not enough to alter its appearance in the telescope but enough to make it seem brighter and enhance the likelihood of its detection. But it is a testament to the vast emptiness of the cosmos that light from even the most distant quasar, emitted when the Universe was in its infancy, has an excellent chance of reaching us without being affected.

The number of galaxy–quasar alignments seen by Webster *et al.*¹ can be explained by standard cosmology, but only if the number of gravitational lenses in and around the galaxies exceeds significantly the number of visible stars. This fits well with the thoughts of most cosmologists, who have been accumulating other evidence for the existence of such 'dark

1. Webster, R. *et al.* *Nature* **336**, 358–359 (1988).
2. Tyson, J. A. *Nature* **334**, 294–295 (1988).
3. Burbidge, G. *Nature* **282**, 451–455 (1979).
4. Arp, H. *Quasars, Redshifts and Controversies* (Interstellar Media, Berkeley, 1987).
5. Stocke, J. T. *et al.* *Astrophys. J.* **315**, L11–L16 (1987).
6. Fugman, W. *Astr. Astrophys.* **204**, 73–80 (1988).
7. Canizares, C. R. *Nature* **291**, 620–624 (1981).
8. Veitri, M. & Ostriker, J. P. *Astrophys. J.* **267**, 488 (1983).
9. Schneider, P. *Astr. Astrophys.* **179**, 80–92 (1987).
10. Kormendy, J. & Knapp, G. R. (eds) *Dark Matter in the Universe* (Reidel, Dordrecht, 1987).

matter' and for the possibility that it is the dominant constituent of the Universe¹⁰. By its very nature, dark matter is elusive and its properties are defined more by speculation than observation. Now, we have the hope of illuminating it with the light beams of distant quasars. The critics of standard cosmology are unlikely to be satisfied — the incidence of lens-induced

alignments is too small to explain their more extreme claims. But the contrast of their almost anecdotal arguments with the precise statistics of the present study further undermines their position. □

Claude R. Canizares is in the Department of Physics and Center for Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

NMR images of microscopic flow

NUCLEAR magnetic resonance (NMR) techniques can be used to measure flow of liquids — in pipes, blood vessels or plants, for example — and NMR imaging is an established tool in clinical diagnosis. But it has not been possible to achieve microscopic resolution by NMR techniques, although this prospect has often been discussed¹⁻⁵. P.T. Callaghan and co-workers have been developing techniques to combine microscopic NMR imaging with measurements of diffusion and flow. On page 399 of this issue⁶ they report the application of their method to measuring water movement within a living wheat grain.

To evaluate the potential of this development, together with other recent applications of NMR imaging of plants and the water movement within them^{7,8}, it should be noted that useful resolution in a microscopic NMR image is usually limited by the signal-to-noise ratio. Other factors are often important, but it is in general possible to modify the imaging techniques so as to approach the limits set by the maximum signal that can be generated by the atomic nuclei in a given volume. The three-dimensional resolution achievable has been variously estimated to be in the range of several to twenty micrometres. Other forms of microscopy can achieve much higher resolution in much shorter image-acquisition times. An NMR image usually requires seconds to hours to obtain, although some techniques can give pictures of somewhat lower quality in tens of milliseconds.

Jenner *et al.* in this issue⁶ illustrate how NMR microscopy can sometimes compensate for its limitations. They measured water movement in the opaque living wheat grain by using the long-known fact that regions in a flowing liquid can be labelled by exciting and observing their nuclear spin populations in ways that cause the signal phases and frequencies to depend on velocity.

The authors use total observation times of almost two hours for the experiments reported here to obtain enough data to reconstruct images, to measure velocities, and to increase the signal-to-noise ratio by averaging the data from repeated observations. They are, however, working at the relatively low frequency of 60 megahertz, corresponding to

a magnetic field of about 1.4 tesla. Commercial spectrometers that operate at 600 megahertz are now available, and imaging accessories can be obtained for spectrometers operating up to 400 megahertz. At these increased frequencies and magnetic fields, measurement times could be reduced by about two orders of magnitude.

Even with greater sensitivity, resolution and speed, the barriers to widespread use of microscopic NMR imaging techniques are formidable. The equipment is large, complex and expensive. To obtain adequate signal-to-noise ratios from small volumes, the radiofrequency-receiving coils must be small and close to the region being imaged, so that the resolution achievable from outside an object is depth-dependent. Although NMR microscopes are not about to become as common as optical or electron microscopes, they do provide the opportunity to carry out carefully chosen investigations that might otherwise not be feasible.

Another consideration, also illustrated by the paper of Jenner *et al.*, is that new techniques can turn up hitherto unsuspected phenomena. The authors believe that the scale of water recirculation within the wheat grain is much greater than previously thought, and it is certainly possible that local transport by bulk flow within plant tissue in general may be a much more active and complex process than previously suspected. NMR methods offer the hope that the effects of factors such as light, water, nutrients, gravity, temperature and infection can be followed over long periods of time in single seeds, in individual plants and in their organs.

Paul C. Lauterbur

1. Lauterbur, P.C. *Nature* **242**, 190–191 (1973).
2. Mansfield, P. & Grannell, P.K. *Physical Rev.* **B12**, 3618–3634 (1975).
3. Eccles, C.D. & Callaghan, P.T. *J. magn. Reson.* **68**, 393 (1986).
4. Lauterbur, P.C. in *NMR in Biology and Medicine* (eds Chien, S. & Ho, C.) 135–140 (Raven, New York, 1986).
5. Aguayo, J.B. *et al.* *Nature* **322**, 190–191 (1986).
6. Jenner, C.F. *et al.* *Nature* **336**, 399–402 (1988).
7. Bottomley, P.A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 87–89 (1986).
8. Johnson, G.A., Brown, J. & Kramer, P.J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2752–2755 (1987).

Paul C. Lauterbur is in the departments of Medical Information Science, Chemistry, Physiology and Biophysics, and Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Illinois 61801, USA.

C. V. Raman on acoustics

THE text below by C. V. Raman and G. A. Sutherland on the Whispering Gallery Phenomenon concludes the short series of excerpts celebrating the centenary of Raman's birth. The phenomenon has fascinated many acoustic scientists and Lord Rayleigh's explanation (*Phil. Mag.* **20**, 1001; 1910; and **27**, 100; 1914) that it is an example of bound-wave propagation is still almost universally quoted as correct. Raman shows that the explanation is incomplete, as the radial fluctuations of sound pressure observed by him are not predicted. At the end of the letter, the authors mention further papers: the most apposite (*Proc. R. Soc. A100*, 424; 1922) includes discussion of the analogous electromagnetic-wave case and elaborates the underlying theory.

J. M. Bowsher, University of Surrey
The very curious and interesting acoustical effects observed in the Whispering Gallery under the dome of St Paul's Cathedral have, as is well known, been explained by the late Lord Rayleigh as due to the curvilinear propagation of sound, the waves which proceed from a source placed close to the wall of the gallery clinging to its surface and creeping tangentially along it. This view was developed mathematically by Lord Rayleigh, the theoretical conclusions arrived at being (a) that the sound-waves travel in a comparatively narrow belt skirting the wall, the thickness of this belt decreasing with the wavelength of the sound; (b) that in this belt the intensity is a maximum near the wall and decreases rapidly and continuously as we proceed radially away from it; and (c) that the intensity does not fluctuate markedly as we proceed circumferentially parallel to the wall.

We have been enabled to carry out an extended series of observations in the gallery with the view of making a precise test of Lord Rayleigh's theory. Our experiments show conclusively that while (a) is substantially accurate, neither of the conclusions (b) and (c) is in accordance with actual facts. Using a steady source of sound placed close to the wall at one point, we found that elsewhere the intensity of the sound showed pronounced oscillations in proceeding inwards radially from the wall, the ear of the observer passing several times through alternate zones of great intensity and of comparative silence. In the latter some of the overtones of the source could be heard clearly, while the fundamental was practically inaudible. These alternations could be demonstrated using a fairly high-pitched source and a sensitive flame as indicator. The distance between the successive zones of silence was about the same as the half-wave-length of the source. There were also distinct periodic fluctuations of intensity in proceeding circumferentially. The latter were not equally distinct in all parts of the gallery, being most marked [opposite] the source.

The circumferential fluctuations of intensity might be interpreted as being due to the stationary interferences of waves passing in opposite directions round the gallery. But the radial fluctuations are less easily explained, and must be regarded as fundamental in any satisfactory theory of the Whispering Gallery. We find that similar effects may be demonstrated in the laboratory with any large circular reflecting surface. We propose at an early opportunity to go more fully elsewhere into the question of the revision necessary in the theory.

From *Nature* **108**, 42 (1921).

Bone physiology

Hypercalcaemia of malignancy

Murray C. Meikle

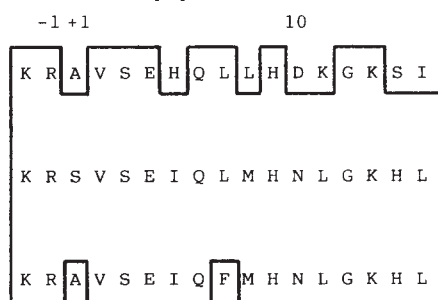
HYPERCALCAEMIA, a complication of malignant diseases, affects up to 40 per cent of patients with certain tumours, the elevation in plasma calcium being caused by tumour products acting on the skeleton to stimulate bone resorption and on the kidney to increase the tubular reabsorption of calcium. Several notable contributions to our understanding of how hypercalcaemia is mediated have been made this year. Probably the most significant development is the recent purification^{1,2} of a parathyroid hormone (PTH)-like peptide from lung and renal carcinomas, whose amino-terminal sequence is very similar to human PTH. And Thompson *et al.* now provide³ direct evidence that this peptide mobilizes bone calcium *in vivo*. These reports are of particular interest as they support a long-standing hypothesis regarding the pathogenesis of tumour-associated hypercalcaemia proposed by Albright⁴ almost 50 years ago.

The gene for human PTH-related protein (PTH-RP) encodes a 177-amino-acid protein consisting of a 36-amino-acid leader sequence and a 141-amino-acid mature peptide^{1,2}. The similarity to PTH, however, is restricted to the amino-terminal region: eight of the first thirteen amino acids of human PTH-RP are identical to human PTH (see figure). Because the amino-terminal region of the human PTH molecule is responsible for its biological activity and binds to the PTH receptor⁵, the similarity of this region seems sufficient to explain the ability of PTH-RP to mimic the effects of PTH. Although PTH-RP is overexpressed by certain tumours, its production is not unique to neoplastic cells. Cultured human keratinocytes also produce PTH-RP (ref. 6), which suggests the peptide may be involved in normal skin physiology and explains why tumours of the squamous type are among those most commonly associated with hypercalcaemia. PTH-RP messenger RNA has also been shown⁷ to be expressed in rat mammary tissue, but only during lactation, apparently as a function of the sucking stimulus. This finding is again noteworthy because hypercalcaemia is a common feature of tumours of the breast.

It seems likely that the genes encoding PTH and PTH-RP arose from a common ancestral gene that became duplicated and separated during mammalian evolution. The PTH gene is located on the short arm of chromosome 11, and the gene encoding PTH-RP has now been mapped⁸ to the short arm of chromosome 12. Chromosomes 11 and 12 are thought to have arisen

by tetraploidization from a single ancestral chromosome⁹, and the short arms of both chromosomes carry other functionally related genes: for example, the genes for lactate dehydrogenases A and B and the Harvey and Kirsten *ras* proto-oncogenes (which are on chromosomes 11 and 12, respectively).

In their new work³, Thompson *et al.* mimicked the constant secretion of the peptide by tumours. They continuously administered peptides, either the first 34



Sequence homologies of the amino-terminal region for human PTH-related peptide (top), human PTH (middle) and bovine PTH (bottom). Eight of the first thirteen amino acids of hPTH-RP are identical to hPTH, which presumably enables the molecule to bind to the PTH receptor: there is little similarity within the remainder of the molecule. (Single-letter amino-acid code.)

residues of bovine PTH, bPTH (1–34), or the equivalent peptide from humans, PTH-RP (1–34), to thyroparathyroidectomized rats fed on a low calcium diet. At an infusion rate of 0.1 nanomoles per hour, both peptides produce hypercalcaemia and nephrocalcinosis and are equally potent in stimulating bone resorption: histomorphometric analysis of tibial long bones reveals a dose-related increase in the number of osteoclast cells. The PTH-RP (1–34) also raises plasma levels of 1,25-dihydroxyvitamin D₃ (1,25(OH)₂vitD₃) in the same animal model⁹. In addition to being a bone-resorbing agent, 1,25(OH)₂vitD₃ also promotes the gastrointestinal absorption of dietary calcium. This finding is likely to be a peculiarity of the animal model, however, because 1,25(OH)₂vitD₃ levels are usually suppressed in patients with tumour-associated hypercalcaemia, probably by elevated plasma calcium levels or impaired renal function¹⁰. An exception appears to be some human lymphotropic virus type-1 lymphomas associated with osteolytic bone lesions and hypercalcaemia, which do produce 1,25(OH)₂vitD₃ (ref. 11).

Evidence is also accumulating to link

the immunoregulatory cytokines interleukin-1 (IL-1) and lymphotoxin (tumour necrosis factor-β) to the hypercalcaemias associated with epithelial carcinomas and haematological malignancies such as myeloma¹². It has been known for some time that IL-1 can stimulate bone resorption *in vitro*¹³. Indeed, IL-1 is the most potent agent yet tested in the standard mouse calvarial bone resorption assay and is effective at concentrations as low as 25 picomolar¹⁴. Experiments conducted on mice by Sabatini *et al.*¹⁵ show that continuous infusions of IL-1 markedly increase plasma calcium and osteoclastic bone resorption. Unlike PTH-RP, however, IL-1-mediated hypercalcaemia appears to be exerted solely through its effect on bone: there is no evidence to suggest that IL-1 exerts any effect on renal tubular function.

Although I have focused here on a few specific examples, it is important to stress that many other factors are capable of stimulating bone resorption, at least *in vitro*, and it is safe to assume that the list will grow. These include tumour necrosis factors α and β, transforming growth factors α and β, epidermal growth factor, platelet-derived growth factor, and arachidonic-acid metabolites such as prostaglandins and leukotrienes. The overexpression of any of these factors by neoplastic cells could theoretically be the cause of osteolysis and hypercalcaemia.

Nevertheless, despite their diverse cellular origin and biological function, all these resorptive agents have a common target cell in bone: osteoblasts and not osteoclasts express the receptors for local and systemic resorptive agents. How osteoblasts transmit resorptive signals to osteoclasts remains unclear, and is the subject of intensive investigation because this relationship has important therapeutic implications, not only for the clinical management of tumour osteolysis, but also for other pathological states characterized by bone loss, such as osteoporosis. □

1. Suva, L. J. *et al.* *Science* **237**, 893–896 (1987).
2. Mangin, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 597–601 (1988).
3. Thompson, D. D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 5673–5677 (1988).
4. Albright, F. *N. Engl. J. Med.* **225**, 789 (1941).
5. Strewler, G. J. *et al.* *J. clin. Invest.* **8**, 803–807 (1987).
6. Meredino, J. J., Insogna, K. L. & Milstone, L. M. *Science* **231**, 388–390 (1986).
7. Thiede, M. A. & Rodan, G. A. *Science* **242**, 278–280 (1988).
8. Comings, D. E. *Nature* **238**, 455–457 (1972).
9. Horiuchi, N. *et al.* *Science* **238**, 1566–1568 (1987).
10. Stewart, A. F. *et al.* *N. Engl. J. Med.* **303**, 1377–1383 (1980).
11. Breslau, N. A. *et al.* *Ann. intern. Med.* **100**, 1–7 (1984).
12. Mundy, G. J. *clin. Invest.* **82**, 1–6 (1988).
13. Gowen, M., Meikle, M. C. & Reynolds, J. J. *Biochim. biophys. Acta* **762**, 471–474 (1984).
14. Heath, J. K. *et al.* *Calcif. Tissue Int.* **37**, 95–97 (1985).
15. Sabatini, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 5235–5239 (1988).

Murray C. Meikle is at the Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 4RN, UK.

Planetary meteorology

Atmospheric oscillations on Mars

Michael Allison

THE thin, carbon dioxide atmosphere of Mars, as revealed by orbiting and lander spacecraft, is a study in exaggerated meteorology, cadenced with remarkably regular annual and daily variations of wind, temperature and pressure, but punctuated with episodic global-scale dust storms. Now, Tillman¹ reports that the further analysis of barometric data from the Viking landers reveals annually synchronized transient oscillations with periods very close to one (martian) day and half a day, with similar events occurring just before two global dust storms. In an accompanying theoretical study, Zurek² interprets the observed transients as either resonant or tidally forced planetary wave modes, excited when the seasonally modulated temperature brings their frequencies into coincidence with the diurnal (daily) and semi-diurnal solar heating.

The dynamic response of an atmosphere to solar driving depends crucially upon its physical and compositional structure as well as the planet's orbital and rotational characteristics. The low density of the martian atmosphere, roughly a hundred times less than that of the Earth, contributes to its exaggerated response to variable forcing. The alignment of the elliptical orbit of the planet with its seasonal solstices produces a hemispheric differential in the evaporation and sublimation of the polar caps, attended by a regular variation in surface pressure from about 700 pascals in the northern summer to 900 in the winter. Data from the Viking meteorological experiments, covering more than three martian years, suggest a peculiarly bimodal behaviour for the northern winter season, one regime characterized by global high-altitude dust storms, the other by intense but regular weather fronts in northern mid-latitudes³.

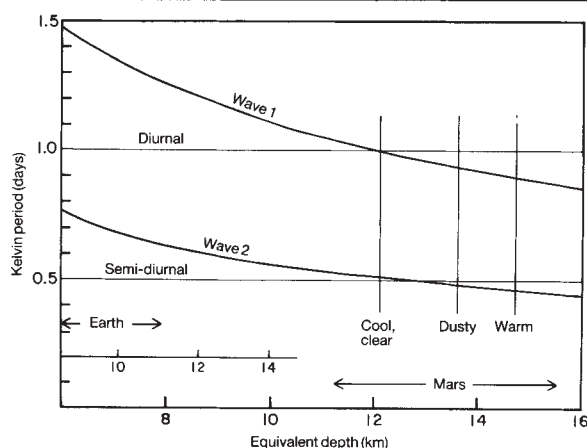
Tillman's new spectral analysis¹ of the Viking pressure data reveals apparently global, nearly diurnal and semi-diurnal transients, persisting for only a few days, which he interprets as normal-mode oscillations of eastward-travelling Kelvin waves. The transient events are observed during the northern mid-winter and recur with surprising precision, synchronized to within a few days of the same season each martian year. In 1977 and 1982 similar transient oscillations were followed by global dust storms, leading Tillman to suggest that the normal modes could be involved in initiating these storms.

In an accompanying theoretical study², Zurek calculates the periods of free and forced wave modes for Mars and the Earth, accounting for the range of vari-

able conditions on both planets. His treatment proceeds according to the methods of classical theory, with a horizontal structure given by the solutions to Laplace's tidal equation⁴, and an assumed separable vertical structure equation, depending on the buoyant stability, radiative cooling, and thermotidal forcing parameters.

The solutions, characterized by the 'equivalent depth' h , for these equations are only marginally different from those for the free modes of the atmosphere (for which h is given simply as $h = H/(1 - R/C_p)$, where H , R and C_p are the pressure scale height, gas constant and specific heat capacity of the atmosphere, respectively).

Kelvin-wave periods in (Earth or Mars) days as a function of the terrestrial and martian 'equivalent depths', for zonal numbers 1 and 2, calculated as solutions to the Laplace tidal equations on a sphere (adapted from Zurek²). Horizontal arrows indicate the span of equivalent depths which depend on temperature and density for which there is a significant response to thermotidal forcing. Vertical lines represent specific values for maximum response to the indicated martian conditions, with generally shorter Kelvin periods for warmer temperatures. The variation of atmospheric temperatures on Mars, as a result of seasonal trends and variable airborne dust loads, probably moves the resonant periods through the diurnal and semi-diurnal harmonics at episodic intervals, thereby stimulating the transients in the Viking lander data now analysed by Tillman². The analogous tidal modes on the Earth are unlikely to be significantly forced because of differences in planetary radius, gravity, temperature and atmospheric composition.



For the longest wavelength Kelvin oscillations (wave-one modes, with one cycle of variation about the equatorial circle), the period on Mars is about 1.1 (martian) days, very close to the natural forcing frequency. Crucially, the period varies inversely as the square root of the mean temperature, decreasing for warmer conditions (see figure). Other types of atmospheric waves all have longer periods and are unlikely to correspond to the observed transients. Also, the equivalent wave-one mode on Earth has a period of 1.5 days, which clearly does not match the forcing period. Although detected in the tropics⁵, its amplitude is much less than that of other modes and noise.

Tillman and Zurek are not the first to suggest the occurrence of high-frequency Kelvin waves on Mars. Hamilton and Garcia⁶ previously proposed a search for these in the Viking pressure data, motivated by their detection of the analogous

events on the Earth. In an earlier study, Zurek⁶ suggested that a diurnal Kelvin mode could be driven by the large-scale structure of the martian topography. And Conrath⁷ found evidence of a similar mode in the analysis of infrared data from the Mariner orbiter spacecraft during the decay of a 1971 dust storm.

Zurek and Leovy⁸ proposed that the observed suppression of the diurnal pressure signal, relative to the semi-diurnal signal, at the Viking landers during dusty conditions could be attributed to the destructive interference of a westward thermal tide and the eastward Kelvin wave. The analysis by Tillman¹ now corroborates the suggestion with a much larger data base than previously available, including a semi-diurnal, supposedly wave-two Kelvin component. The reported synchronization of the transients to within a few solar days each martian year, with nearly simultaneous readings at both

landers, and the coincidence of similar events with global dust storms, could reflect a finely tuned coupling to special conditions in the planet's global meteorology.

Tillman proposes that the transients are purely free modes resonantly amplified by variations in the near-surface temperatures and static stability. But Zurek's analysis² indicates that the atmospheric temperature data have the wrong seasonal trend for resonance alone to account for these events near the time of the global pressure minimum. The real situation, as he suggests, probably involves an optimal convolution of the atmospheric resonance

1. Tillman, J.E. *J. geophys. Res.* **93**, 9433-9451 (1988).
2. Zurek, R.W. *J. geophys. Res.* **93**, 9452-9462 (1988).
3. Leovy, C.B. et al. in *Recent Advances in Planetary Meteorology* (ed. Hunt, G.E.) 69-84 (Cambridge Univ. Press, 1985).
4. Longuet-Higgins, M.S. *Phil. Trans. R. Soc. A* **262**, 511 (1968).
5. Hamilton, K. & Garcia, R.R. *J. geophys. Res.* **91**, 11867-11875 (1986).
6. Zurek, R.W. *J. Atmos. Sci.* **33**, 321-337 (1976).
7. Conrath, B.J. *J. Atmos. Sci.* **33**, 2430-2439 (1976).
8. Zurek, R.W. & Leovy, C.B. *Science* **213**, 437-439 (1981).

with thermotidal forcing, modulated by rapid changes in the atmospheric dust load.

It still is not clear why the otherwise regular seasonal variations on Mars are differentiated by the occurrence of planet-encircling dust storms in some years which in others are replaced by alternating synoptic weather fronts. Leovy *et al.*³ offer the fascinating conjecture that the two-regime behaviour is a meteorological example of bifurcation (forking), with alternate branches selected by small random

fluctuations in the coupling between dust, atmosphere and the polar caps during the preceding seasons. Further study of the Viking data and further space missions to Mars are needed to resolve the riddle. Whatever the answer, it is interesting to have found what could be a branch point in atmospheric chaos signalled by the clearly tuned ring of an almost normal mode. □

Michael Allison is in the NASA Goddard Institute for Space Studies, 2880 Broadway, New York, New York 10025, USA.

Forest ecology

Blow, blow thou winter wind

Peter D. Moore

WINDS of hurricane force (with wind speeds in excess of 33.5 metres per second) are rare in the British Isles, but the storms of last year raised important ecological questions. Of particular interest is the influence of such rare catastrophes on natural woodlands. New studies of forest recovery in North America by Foster^{1,2} are pointing to the important role of high winds in temperate forests; they could be instrumental in the creation and maintenance of mosaic patterns and hence the diversity of these woods.

The frequency and strength of wind in the recent history of the Earth's climate has been studied by extrapolation from numerical modelling³, but more direct information on wind history is also available from fossil sources and from the study of historically documented wind damage to vegetation. In their reconstructions of the history of winds, Hobgood and Cerveny³ conclude that during the last glacial maximum (about 18,000 years ago) tropical storms generating winds of hurricane force were scarcer, less intense and shorter than those of the present day. As a consequence, less precipitation was brought into some warm temperate areas, such as Florida, which remained fairly arid. But occasional high winds have nevertheless been a feature of the temperate climate throughout the past 10,000 years and could have influenced the development, structure and composition of the migrating and reassembling forests of the mid- and higher latitudes.

Direct evidence of the effects of wind on forests is hard to come by. Perhaps the most convincing comes from the fossil record — tree-trunk alignment in the peaty remains of swamp forests, submerged and preserved beneath the rising sea-levels of coastal lowlands in post-glacial times. A 6,000-year-old fossil forest in mid-Wales (see figure) has been sufficiently well preserved and is adequately exposed at low tide for Taylor to analyse⁴ the orientation of 160 fallen tree trunks. There is a strong tendency for

alignment along an approximately south-west to north-east axis. It is not known whether individual trees died as a result of a single catastrophic storm or of a strong prevailing wind.

In New England, Foster^{1,2} studied directly the effects of a hurricane which in 1938 devastated many forests in the area. The storm was accompanied by heavy rain, generated winds in excess of 55 metres per second and caused damage to forests along a 100-kilometre-wide track northwards through central New England. In the study site, Harvard forest,



Buried treasure — submerged forest peat at Borth, mid-Wales, contains well-preserved fossils.

70 per cent of the trees had been torn down. Foster shows that the damage is positively related to the age of each stand, and that some species, such as white pine (*Pinus strobus*), are more susceptible to damage than others, for instance hardwood species. Wind susceptibility, however, may vary with season, for trees in full leaf with roots in wet soils are more likely to suffer wind damage.

The regular orientation of fallen timber at Harvard forest is reminiscent of the fossil data from Wales and lends support to the conclusion that the fossil forest was influenced by wind. But the subsequent recovery of the damaged forest in New England has left a mosaic of successional stages, depending on the local physiography and the original vegetation cover.

In the Harvard forest, most trees had been uprooted rather than broken off,

which could have been a consequence of the accompanying rainfall and the poorer anchorage afforded by waterlogged soils. The same is true of the British storm last year, which happened after several weeks of very wet weather. Thomas Hardy, in his account⁵ of tree planting in the nineteenth century, describes how the roots of saplings were spread towards the south-west to provide subsequent strength in just such times of gales. So Victorian foresters in Britain were evidently aware of the dangers of uprooting by wind.

Uprooting, rather than the breakage of tree boles, could be significant for the pattern of subsequent regeneration. In a piece of experimental forest destruction, Collins and Pickett⁶ created gaps in hardwood forest in Pennsylvania by cutting canopy and understorey trees at 1 metre above the ground. Herb layer vegetation shows very little response to such gaps. It is reasonable to suppose that soil disturbance is needed if there is to be additional recruitment from the seed bank. Uprooting of trees by wind would be expected to generate a more diverse response in the vegetation, for it would result not only in the creation of a greater range of light micro-environments but also the stimulation of dormant seeds assembled at earlier successional stages.

A woodland affected by periodic winds will therefore consist of a mosaic of

recovery stages, micro-successions within the forest that add variety to the range of habitats available to plants and animals within the system. There is a catastrophic storm in New England about once every 100–150 years. In Britain, it may be 300 years since a storm like that of 1987 was experienced. But in natural forest ecosystems, such an infrequent event may be very significant in determining the structure and composition of the canopy and perhaps even the diversity of the ground flora. □

1. Foster, D.R. *J. Ecol.* **76**, 105–134 (1988).

2. Foster, D.R. *J. Ecol.* **76**, 135–151 (1988).

3. Hobgood, J.S. & Cerveny, R.S. *Nature* **333**, 243–245 (1988).

4. Taylor, J.A. *Prog. Geogr.* **5**, 247–334 (1974).

5. Hardy, T. *The Woodlanders* (Macmillan, London, 1974).

6. Collins, B.S. & Pickett, S.T.A. *J. Ecol.* **76**, 437–450 (1988).

Peter D. Moore is Reader in Ecology in the Division of Biosphere Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

Magnetic soils

Magnetite *sans* microbes

Subir K. Banerjee

THE uppermost layers of soil are commonly magnetized many times more strongly than the layers deeper in the soil profile or the country rock upon which the soil has developed. Maher and Taylor report on page 368 of this issue¹ the successful extraction of single microcrystals of pure magnetic fractions from two types of British soil. They show convincing, if circumstantial, evidence that the greater magnetization arises as the soil is formed in a process that produces impurity-free magnetite microcrystals with perfect crystal faces which are only 10–30 nm across, and are concentrated in the topsoil.



A chain of biogenic magnetite grains identified by Petersen et al. from deep-sea sediment cores. But not all magnetite is biogenic, Maher and Taylor show¹. (Scale bar, 0.1 μ m; from ref. 8.)

This observation is important not only for studies of rock magnetism, but also for biomagnetic studies because, since the discovery^{2,3} of similar perfect microcrystals of magnetite in freshwater bacteria *Aquaspirillum magnetotacticum*, there has been a general belief that such magnetite is produced only by biologically mediated processes.

Following the discovery of biogenic magnetite, high-quality samples of magnetite were produced in laboratory bacterial cultures and recovered from lakes, bogs and deltas. A whole volume has recently been written on such magnetite biomineralization in bacteria as well as in higher animals such as bees, birds and mammals⁴. A kind of biomagnetic bandwagon has arisen, which alarms me, by which bacterial origins have often been proposed, in the absence of direct bacterial evidence, for magnetic particles in ancient iron formations⁵, hydrocarbon deposits⁶ and in suboxic marine sediments^{6,7}. Clear evidence of biogenic magnetite in ancient marine sediments has been found once only⁸ (see figure) despite a long, hard search. The new work of Maher and Taylor in this issue¹ should halt, or at least slow, this bandwagon, because if perfect magnetite crystals can be produced inorganically in soil their presence in lake,

river and ocean sediments, whether recent or ancient, need not be attributed to biogenesis. Then Earth scientists could use such crystals as tracers of past erosional, hydrological and atmospheric regimes, as they have in the past⁹.

In proving their thesis, Maher and Taylor had to reject three alternatives. First, that the magnetic mineral was maghemite, which is similar to magnetite. Maghemite is found in burnt soil and is formed by decomposition of aluminium- or titanium-bearing precursors like haematite or goethite. Energy-dispersive X-ray analysis showed that maghemite was not present, as there were only trace amounts of impurities, zinc and manganese.

Second, the authors had to show the crystals were not a remnant of the country rock. Such detrital magnetite had previously been found from magnetic¹⁰ and Mössbauer studies¹¹. The authors rely on their own synthesis¹² of ultrafine magnetite under soil-forming (near-neutral-pH) conditions at room temperature. These synthetic magnetite grains are morphologically and magnetically similar to the magnetic extracts from soil.

Lastly, the authors had to exclude the possibility of a biogenic origin. Because the size, shape, purity and magnetic properties of the magnetic extract are similar to those of bacterially produced magnetite, how can one distinguish between biogenic and pedogenic origins? Maher and Taylor are remarkably frank in saying that on the basis of shape, purity and magnetic properties alone, the two origins are indistinguishable. But their soil magnetite does not include associated bacterial cells and a realistic laboratory simulation of pedogenesis does form very similar microcrystals of magnetite, making this the simplest explanation. □

1. Maher, B.A. & Taylor, R.M. *Nature* **336**, 368–370 (1988).
2. Blakemore, R. *Science* **190**, 377–379 (1975).
3. Frankel, R.B., Blakemore, R.P. & Wolfe, R.S. *Science* **203**, 1355–1356 (1979).
4. Kirschvink, J.L., Jones, D.S. & McFadden, B.J. (eds) *Magnetite Biomineralization and Magnetoreception in Organisms* (Plenum, New York, 1985).
5. Elmore, R.D. et al. *Nature* **325**, 428–430 (1987).
6. Lovley, D.R., Stolz, J.F., Nord, G.L. Jr & Phillips, E.J.P. *Nature* **330**, 252–254 (1987).
7. Kirschvink, J.L. *Earth planet. Sci. Lett.* **59**, 388–392 (1982).
8. Petersen, N., von Dobeneck, T. & Vali, H. *Nature* **320**, 611–615 (1986).
9. Thompson, R. & Oldfield, F. *Environmental Magnetism* (Allen & Unwin, London, 1986).
10. Uzoemir, U. & Banerjee, S.K. *Earth planet. Sci. Lett.* **59**, 393–403 (1982).
11. Longworth, G. et al. *J. Soil Sci.* **30**, 93–110 (1979).
12. Taylor, R.M., Maher, B.A. & Self, P.G. *Clay Min.* **22**, 411–422 (1987).

Subir K. Banerjee is a professor of Geology and Geophysics at the University of Minnesota, Minneapolis, Minnesota 55455, USA.

Daedalus

Scale model

FISH swim much more efficiently than submarines. One explanation invokes the flexible skin of the fish, which yields to local water-pressure fluctuations and damps out turbulence before it can build up. The wasteful eddies which grow around a rigid hull and stream away from it, carrying energy with them, are thus stifled at birth.

And this, says Daedalus, is why fish have scales. Each scale can yield independently to its local pressure variations, without being constrained by the different motions of its neighbours. To test his theory, Daedalus has constructed an artificial fish, with scales attached to its metal body on highly-damped polymeric mountings. Various configurations of this model are being tested in DREADCO's towing tank, to discover what size, distribution and damping of the scales gives the lowest drag. If this optimum closely mimics the scale pattern of real fish, Daedalus's theory will be confirmed. But even if it does not, like all good scientists he has a second theory up his sleeve. Fish scales may not be passive dampers at all, but active ones.

An active eddy-damping scale would sense the pressure perturbation of an incipient eddy, and then move in such a way as to cancel it perfectly. Daedalus's Mark II artificial fish will carry each of its scales on a piezoelectric element governed by a local control chip. Any pressure fluctuation on the scale elicits a pre-programmed response. DREADCO's hydrodynamicists are devising local scale strategies which should damp out eddies completely, imposing pure laminar flow around the 'fish' and giving it the lowest possible drag.

All this, says Daedalus, may only be mimicking nature. But he goes further: he wants his active-scale skin to *propel* his fish. Under central computer control, the scales will generate peristaltic ripple patterns running from its head to its tail. Imposed on the whole wetted surface of the fish, even a modest ripple pattern should produce enough reaction to propel it. A suitably-programmed active skin could generate more complex peristaltic patterns, to stabilize and steer the fish as well. This elegant integrated marine-propulsion system deserves to be scaled up. An active-skin ship, coupled to the water over its whole hull and driven silently in pure laminar flow, would be far more efficient than our present rigid, turbulent, propeller-driven monsters. It could manoeuvre in any direction, even sideways, indeed even upwards, to reduce its draught if it went aground. And barnacles and similar fouling organisms would be numbed and dislodged by the ceaseless rapid vibration of its scales.

David Jones

Understanding the origins of AIDS viruses

SIR—Understanding the origins of the AIDS viruses is of obvious importance. Unfortunately, recent suggestions regarding the evolutionary history of HIV-1, and its relationship to the two other so-far elucidated groups of primate lentiviruses (HIV-2/SIV_{MAC} and SIV_{AGM}) are astonishingly divergent.

At one extreme, Fukasawa *et al.* suggest that HIV-1 and SIV_{AGM} diverged “gradually in concert with the evolution of primates”, implying that the two viruses split millions of years ago. At the other, Smith *et al.*² conclude that HIV-1 and HIV-2 may have evolved from a common ancestor “as recently as 40 years ago”. The difference between these two dates is remarkable as HIV-1, HIV-2 and SIV_{AGM} are approximately equidistantly related to one another (see below).

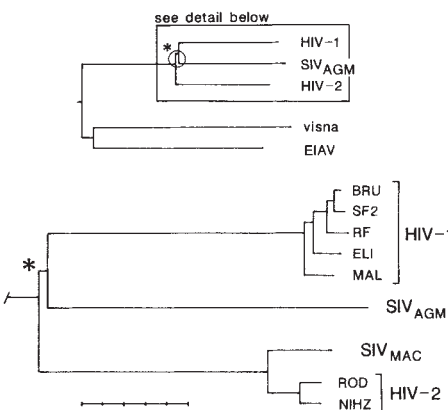
In fact, both dates seem unlikely in the light of present knowledge. First, nucleotide substitutions in retroviral (RNA) genomes accumulate about a million times faster than in the nuclear DNA of higher organisms^{3,4}. Because HIV-1 and SIV_{AGM} are only 20 times more divergent than human and Old World monkey DNA sequences⁵, it is likely that these two viruses have been separated for much less than 1,000 years. Second, serological evidence suggests⁶ that HIV-1 existed in central Africa about 30 years ago, before 1960, and our recent analysis⁴ of diverse HIV-1 isolates suggests that their common ancestry goes back to about that time. As HIV-1 and HIV-2 are far more divergent than any two HIV-1 isolates^{2,7}, it is very unlikely that HIV-1 and HIV-2 diverged only 40 years ago. We estimate that HIV-1, HIV-2 and SIV_{AGM} all diverged about 150 years ago.

To study the evolutionary history of these rapidly evolving viruses, we examined the more conserved parts of the genome. The *pol* gene is best conserved, and phylogenetic analysis using non-primate lentiviruses (such as visna) as outgroups reveals the likely relationships among the human and simian viruses (see figure). This phylogeny has two facets of immediate interest: the three main groups of primate lentiviruses, HIV-1, HIV-2 and SIV_{AGM}, diverged at about the same time (asterisk in the figure); and the branch lengths, indicating the degrees of divergence from that point, are similar down each descendant lineage, providing evidence for similar rates of molecular evolution among different groups. The latter result contradicts the suggestion⁸ that a more virulent virus, such as HIV-1, may evolve considerably more rapidly than less virulent ones, such as HIV-2 and SIV_{AGM}.

Because there seems to be a good molecular clock for HIV, it can be calibrated by placing only one time point on

the phylogeny, to estimate how long ago the divergence at the asterisk took place. We have recently done this⁴ for the divergence times among HIV-1 isolates and obtained rates of evolution yielding divergence time estimates consistent with epidemiological data. HIV-1 ELI, isolated in 1983, was estimated to have diverged from the non-African isolates BRU, SF2 and RF around 1969. By extrapolation, the divergence of HIV-1, HIV-2 and SIV_{AGM} is estimated to have occurred between 140 and 160 years ago. Yokoyama *et al.*⁷ estimated that HIV-1 and HIV-2 diverged about 280 years ago, on the assumption that the lentivirus *pol* genes have evolved at the same rate as viral oncogenes, that is 0.5×10^{-3} substitutions per (nonsynonymous) site per year³. Our rate estimate for the HIV-1 *pol* gene is higher (0.96×10^{-3} in the conserved regions considered here), accounting for our shorter estimate of the divergence time.

Because HIV-1 and HIV-2 seem to have evolved at similar rates, we estimate that HIV-2 and SIV_{MAC} diverged only about 30 years ago. Phylogenetic analysis⁷ of a wider group of retroviruses supports the position of the root of the upper tree as indicated in the figure, and so suggests that the rate of evolution in the visna lineage has been similar to that in the



Phylogenetic relationships among lentiviruses. HIV, Human immunodeficiency virus; SIV, simian immunodeficiency virus (AGM, African green monkey; MAC, macaque); visna, ovine lentivirus; EIAV, equine infectious anaemia virus. Horizontal distances, proportional to the number of nucleotide substitutions along each branch; vertical separation for clarity only. Distances among viruses are in nonsynonymous nucleotide substitutions⁹ among *pol* genes (note that areas of the gene in which alignment is difficult were excluded, leaving 506 codons for comparison). Total length of scale bar, 0.05 substitutions per site. Phylogeny constructed using the neighbour-joining method¹⁰, which sequentially clusters sequences by minimizing the total branch length of the tree. The same branching order was obtained by the transformed distance method¹¹ using as references visna and EIAV.

primate lentiviruses. Therefore, we can tentatively estimate that the latter separated from visna about 300 years ago.

Finally, note that both HIV-1 and HIV-2 cause AIDS, whereas SIV_{AGM} does not¹. Our analysis suggests that all three viruses diverged at about the same time, suggesting that all three viruses were originally pathogenic, with SIV_{AGM} having lost its pathogenicity after the divergence. Alternatively, if the three viruses were originally nonpathogenic, then HIV-1 and HIV-2 must have acquired their pathogenicity independently after divergence.

PAUL M. SHARP

Department of Genetics,
Trinity College, Dublin 2, Ireland

WEN-HSIUNG LI

Center for Demographic and Population
Genetics,
University of Texas,
PO Box 20334,
Houston, Texas 77225, USA

1. Fukasawa, M. *et al.* *Nature* **333**, 457–461 (1988).
2. Smith, T.F., Srinivasan, A., Schochetman, G., Marcus, M. & Myers, G. *Nature* **333**, 573–575 (1988).
3. Gojobori, T. & Yokoyama, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4198–4202 (1985).
4. Li, W.-H., Tanimura, M. & Sharp, P.M. *Molec. biol. Evol.* **5**, 313–330 (1988).
5. Li, W.-H., Tanimura, M. & Sharp, P.M. *J. molec. Evol.* **25**, 330–342 (1987).
6. Nahmias, A.J. *et al.* *Lancet* **i**, 1279–1280 (1986).
7. Yokoyama, S., Chung, L. & Gojobori, T. *Molec. biol. Evol.* **5**, 237–251 (1988).
8. Penny, D. *Nature* **333**, 494–495 (1988).
9. Li, W.-H., Wu, C.-I. & Luo, C.-C. *Molec. biol. Evol.* **2**, 150–174 (1985).
10. Saitou, N. & Nei, M. *Molec. biol. Evol.* **4**, 406–425 (1987).
11. Li, W.-H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1085–1089 (1981).

How to make models for behaviour of clouds

In his commentary “Jumping the greenhouse gun”, John Maddox states that “The well-known weakness of the climatic models, in this connection, is that real clouds (as distinct from average cloudiness) are on the face of it a source of negative feedback that may substantially moderate the expected size of the increase in temperature of the surface of the Earth.” Here I explain why the sign of cloud feedback is very difficult to estimate by intuition and why, therefore, it is not surprising that recent general circulation model simulations with predicted (non-average) clouds give positive cloud feedbacks. To do this it is necessary to develop the following framework for the analysis of the feedbacks in the climatic system.

The net radiation at the top of the Earth’s atmosphere can be expressed as $N = N(E, T, I)$, where E is a vector of quantities ‘external’ to the climate system (for example, the CO₂ concentration), T is the surface temperature, and I is a vector of quantities ‘internal’ to the climate system (for example, clouds). A small change in the energy flux, ΔN , can be expressed as $\Delta N = \Delta Q - (G_0^{-1} - F) \Delta T$. Here ΔQ is the change in N due to a

change in one or more external quantity; $-G_o^{-1} \Delta T$ is the change in N due to the change in T alone, where $G_o = -(\partial N / \partial T)^{-1} \approx 0.3^\circ \text{C} (\text{Wm}^{-2})^{-1}$ is the gain of the climate system without feedback; and $F \Delta T$ is the change in N due to the change in the internal variables I through their dependence on T (ref. 2). When the equilibrium ΔT is reached in response to the forcing ΔQ , $\Delta N = 0$ and $\Delta T = G_o \Delta Q / (1 - f)$, where f is the sum of the individual feedbacks $f_j = G_o \partial N / \partial I_j dI_j / dT$. Thus the feedback of a physical process j depends on three quantities: (1) the zero-feedback gain of the climate system G_o , (2) the sensitivity of the net flux N to the process as measured by $\partial N / \partial I_j$, and (3) the sensitivity of the process to the surface temperature as measured by dI_j / dT .

There are two possible feedbacks due to greenhouse-gas-induced changes in clouds. One is cloud-cover feedback, $f_c = G_o \delta dC/dT$, where C is cloud cover and $\delta = \partial N / \partial C$ is the change in N resulting from a change in C . The second is cloud-optical-depth feedback, $f_\tau = G_o \phi d\tau/dT$, where τ is the cloud optical depth and $\phi = \partial N / \partial \tau$ is the change in N resulting from a change in τ . Several radiative transfer model studies (summarized in ref. 2) have shown that δ is negative for low and middle clouds, and is positive for high thin cirrus clouds. Other radiative transfer model studies (also summarized in ref. 2) have shown that ϕ is also negative for low and middle clouds, and is positive for high thin cirrus clouds. These results can be deduced from intuition. However, δ and ϕ are not the cloud-cover and cloud-optical/depth feedbacks. To determine the latter requires knowledge also of dC/dT and $d\tau/dT$, the changes in cloud cover

and cloud optical depth which occur as a result of changes in surface temperature. In fact, it is the vertical integrals throughout the atmosphere of $\delta dC/dT$ and $\phi d\tau/dT$ which respectively determine the signs of the cloud-cover and cloud-optical-depth feedbacks. It is these vertical integrals which are virtually impossible to estimate by intuition.

Consequently, it is not surprising that the cloud-cover feedback in recent general circulation model simulations of CO_2 -induced equilibrium climate change is positive as a result of a decrease in low and middle cloud cover ($\delta < 0$ and $dC/dT < 0$) and an increase in high cloud cover ($\delta > 0$ and $dC/dT > 0$)³⁻⁵. Similarly, it is also not surprising that the cloud optical depth feedback in a recent general circulation model simulation for a 2% increase in solar constant is positive due to an increase in the optical depth of high thin clouds ($\phi > 0$ and $d\tau/dT > 0$)⁶⁻⁸.

MICHAEL E. SCHLESINGER

Department of Atmospheric Sciences
and Climatic Research Institute,
Oregon State University,
Corvallis, Oregon 97331, USA

1. Maddox, J. *Nature* **334**, 9 (1988).
2. Schlesinger, M.E. in *The Potential Climatic Effects of Increasing Carbon Dioxide* (eds MacCracken, M.C. & Luther, F.M.) 280-319 (NTIS, Springfield, Virginia, 1985); in *Physically-Based Modelling and Simulation of Climate and Climatic Change* (ed. Schlesinger, M.E.) 653-736 (Reidel, Dordrecht, 1988).
3. Hansen et al. in *Climate Processes and Climate Sensitivity* (eds Hansen, J.E. & Takahashi, T.) 130-163 (American Geophysical Union, Washington, DC, 1984).
4. Wetherald, R.T. & Manabe, S. *J. Atmos. Sci.* **45**, 1397-1415 (1988).
5. Schlesinger, M.E. in *Contribution of Geophysics to Climate Change Studies, Proc. IUGG Symp. 15* (eds Berger, A. et al.) (Am. Geophys. Un., in the press).
6. Roeckner, E., Schlese, U., Biercamp, J. & Loewe, P. *Nature* **329**, 138-140 (1987).
7. Schlesinger, M.E. *Nature* **333**, 303 (1988).
8. Roeckner, E. et al. *Nature* **333**, 303 (1988).

Unusual segregation of cystic fibrosis alleles

SIR—In a previous scientific correspondence¹, we described an unusual segregation of the cystic fibrosis (CF) allele to males. This was based on the observation of 60 French siblings of CF patients. Subsequently, two other small studies^{2,3} failed to confirm the initial claim.

We have now collected data from 21 European groups (including those in refs 2 and 3) and six North American groups, in which the CF status was analysed using the probes XV-2c and CS.7 (ref. 4), KM19 (ref 5), J3.11 (ref. 6) and *MET* (ref. 7). For each probe the alleles that are linked to the cystic fibrosis mutation were determined using at least one index case; no crossover families were used in this study.

In total, 1,003 unaffected siblings of CF probands were investigated. The unaffected siblings, have a male/female ratio of 1.018

Normal homozygotes, heterozygotes with the CF gene of paternal origin, and heterozygotes with the CF gene of maternal origin each represent approximately one third of the total number of unaffected children, as expected.

As shown in the table, of 349 normal homozygotes, 201 are girls and 148 are boys. This difference is statistically significant ($\chi^2 = 8.05$, $P < 0.01$). Of 654 heterozygotes there are 358 boys and 296 girls. This difference is also significant ($\chi^2 = 5.87$, $P < 0.02$). For neither sons nor

daughters who are heterozygotes, is there a statistically significant difference in the likelihood that the CF gene is of maternal or paternal origin.

For heterozygous offspring, there are more males than females, but this difference is not statistically significant whether the CF gene is inherited from the father ($\chi^2 = 3.55$) or the mother (2.39). The sex ratio in heterozygotes inheriting the gene from the mother or the father is the same.

Although the statistical significance regarding preferential transmission of the CF gene from a parent to a child can be calculated in a variety of ways, the expanding data do not support our original conclusion¹ that the CF gene is preferentially transmitted from fathers to sons. But we conclude that there is a distortion in the expected sex ratio for normal homozygotes which results in an excess of girls, and a reciprocal difference in the sex ratio of heterozygotes which results in an excess of boys inheriting the CF gene from either their father or their mother.

The mechanism underlying the distortion in the sex ratio remains unclear. It could operate before fertilization, at the time of fertilization or through selective survival after fertilization. It will be important to determine if there is any distortion in the sex ratio of homozygous affected individuals at conception, although current postnatal detection does not provide evidence for a large excess of affected males. The data do not offer an immediate explanation for the high incidence of CF in the population but they could be of importance in this respect.

A. KITZIS ET AL. *

Institut de Pathologie Moléculaire,
CHU Cochin, 75014 Paris, France

*J.C. Chomel, A. Haliassos, L. Tesson, J.C. Kaplan, J. Feingold (Paris), G. Giraud, A. Labbe, B. Dastugue (Clermont-Ferrand), V. Dumur, J.P. Farriaux, P. Roussel (Lille), C. Ferrec (Brest), M. Vidaud, M. Goossens (Créteil), D. Bozon, M. Auvinet, V. Chambon, J. Andre (Lyon), W. Lissens, M. Bonduelle, I. Liebaers, P. Cochaux, G. Vassart (Brussels), P. Willems (Antwerp), G. Duckworth-Raysieck, B.S. Kerem, L.C. Tsui, P.N. Ray (Toronto), M. Krawczak, J. Schmidtke (Göttingen), G. Novelli, B. Dallapiccola (Rome and Urbino), G. Gasparini, P.F. Pignatti (Verona), M. Seia, M. Ferrari (Milan), M. Devoto, G. Romeo (Geneva), M. Schwarz, M. Super, A. Ivanson, A.P. Read (Manchester), L. Meredith (Cardiff), A. Curtis (Edinburgh), R. Williamson (London), A.L. Beaudet, G.L. Feldman, W.E. O'Brien (Houston), A.M. Bowcock, L.L. Cavalli-Sforza (Stanford), F. Gilbert (New-York), J. Brame (La Jolla), M.C. King (Berkeley).

1. Kitzis, A. et al. *Nature* **333**, 215 (1988).
2. Ten Kate, L.P., Te Meerman, G.J., Buys, C.H.C.M., Halley, D.J.J. & Oostra, B. *Nature* **334**, 20 (1988).
3. Tümmeler, B., Aschendorf, A., Darnedde, T. & Hundrieser, J. *Nature* **334**, 110 (1988).
4. Estivill, X. et al. *Nature* **326**, 840-845 (1987).
5. Estivill, X. et al. *Genomics* **1**, 257-263 (1987).
6. Wainwright, B.J. et al. *Nature* **318**, 384-385 (1985).
7. White, R. et al. *Nature* **318**, 382-384 (1985).

Insect eating in Japan

SIR—In response to John Treherne's review of Vincent Holt's book *Why Not Eat Insects?* (*Nature* **335**, 126; 1988), I wish to provide some information on insectivory in Japan.

In autumn, in agricultural villages, rice hoppers (*inago*) are collected from the

	Heterozygous		Homozygous		
	Maternal CF origin	Paternal CF origin	Total	Normal	Total
Boys	178	180	358	148	506
Girls	150	146	296	201	497
Total	328	326	654	349	1003

farms and after the removal of wings and limbs are cooked with soy sauce and sugar, and served. The merits of the practice are twofold: first, the use of a good source of animal protein, particularly notable in a country which until a century ago detested killing land mammals; second, as Treherne pointed out, removal of an agricultural pest.

Processed rice hoppers (mainly from Nagano prefecture) called *tsukudani* are sold in packets in supermarkets and are mainly eaten as a snack with alcoholic beverages. In addition to rice hoppers, wasp maggots (*hachinoko*) are also eaten in Japan. The National Food Composition Tables (Resources Council, *Standard Tables of Food Composition in Japan*, 4th edn; Science and Technology Agency, Tokyo, 1982) provide the nutritional composition of *tsukudani* from *Oxya yezoensis* Shiraki and canned maggot of the wasp *Vespa japonica* Saussure. Their protein compositions are 22.5 and 15.7 grams per 100 grams of edible portion, respectively.

SACHI SRI KANTHA

Laboratory of Marine Biochemistry,
Faculty of Agriculture,
University of Tokyo,
Bunkyo-ku, Tokyo 113, Japan

Latex gloves off in virus porosity dispute

SIR—Two of the authors of the recent letter¹ about latex gloves and viruses, are at Advanced Biotechnologies Inc., which may explain their ignoring a century of microbiology. Pasteur showed that broth remains sterile in a flask stoppered with an open glass tube if the tube has two to three bends. Particles hate labyrinths and get stuck in the first or second corner. Sterilization by filtration through candles or membranes was always based on the principle that pores may be much larger than the microbe to be stopped.

Arnold *et al.* show a nice electron-micrograph with channels and pores in the latex glove but provide no evidence that viruses pass through them. Contrary results² are unjustly rejected. The use of several gloves is certainly interesting for the glove industry but may reduce the handling skill of operators.

JOSEPH HUPPERT
GEORGE MATHE

Institut de Cancerologie et
Immunogénétique,
94800 Villejuif, France

SIR—Arnold *et al.*¹ report that latex gloves have pits on both interior and exterior surfaces as judged by scanning electron microscopy. They compare their findings to ours² and conclude that double gloving, possibly supplemented by surfactant virucide, would be prudent for those handling material infected by HIV or hepatitis B virus.

Our study used reputable gloves avail-

able to laboratories and hospitals, and a full description and source of all the gloves was given. Although we did not look for cavities by electron microscopy, we reported that even in the event of a breach of the latex rubber the virus in contact with latex seems to be inactivated.

None of the gloves tested by Arnold *et al.* are listed in their letter and one of the authors (C.H. Fox) has led us to believe that the gloves used were known not to be of high quality relative to medical applications. It is very important that the type of glove, the type of latex and the processes used to make the latex are taken into consideration, especially before reporting what could be construed as alarming results.

A.G. DALGLEISH
MIREK MALKOVSKY

Clinical Research Centre,
Harlow, Middlesex HA1 3UJ, UK

SIR—We wish to correct the impression given by Arnold *et al.*¹ that latex gloves contain micronized fissures or channels impairing their integrity.

We have been examining latex gloves, condoms and other latex products for many years, using transmission and scanning electron microscopy as well as energy-dispersive X-ray analysis. The surfaces of latex films may sometimes exhibit pits or protuberances, and cross-sections of films may reveal residual particle interfaces, but the presence of "ducts" or "channels" has never been observed even at moderately high magnifications.

In our view the "5-micron tortuous channels" reported by Arnold *et al.* are probably artefacts of sample preparation. It is usual, in preparation for electron microscopy, to apply an electrically conductive coating to polymers. The coating may be of carbon, gold or gold-palladium alloy, or combinations of these.

Unless precautions are taken during the coating process, and subsequent examination of specimens, it is possible to produce artefacts which strongly resemble the "channels" shown in the published micrograph.

We conclude that Arnold *et al.* were misled by cracking of the coating applied during sample preparation. Their recommendation that two pairs of gloves are required to ensure adequate protection is therefore unnecessary.

T. D. PENDLE
A. J. COBBOLD

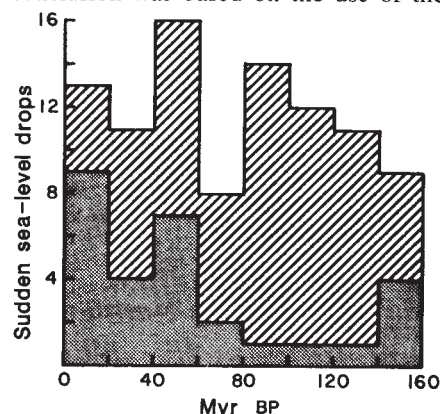
Tun Abdul Razak Laboratory,
Brickendonbury,
Hertford SG13 8NL, UK

1. Arnold, S.G., Whitman, J.E. Jr, Fox, C.H. & Cottler-Fox, M.H. *Nature* **335**, 19 (1988).
2. Dalgleish, A.G. & Malkovsky, M. *Brit. J. Surg.* **76**, 171–172 (1988).

■ Two of the authors of the letter "Latex gloves not enough to exclude viruses" have since moved. The present address for S.G. Arnold and J.E. Whitman Jr is: Advanced Biotechnologies Inc., Rivers Park II, 9108 Guilford Road, Columbia, Maryland 21046, USA.

Magnetic reversal rate and global sea level trends

SIR—Morris and Muller¹ recently commented on Gaffin's² finding of a negative correlation between the long-term trend in global sea level over the past 150 Myr—the so-called first-order sea-level curve of Vail *et al.*³—and the frequency of geomagnetic reversals. In order to support their hypothesis that reversals are caused by sudden climate coolings and drops in sea level⁴, however, they persist in the erroneous conclusion of their earlier paper, that the period of generally high sea level during the Cretaceous, from about 80 to 120 Myr BP, was also a time of a reduced number of shorter term (second-order) sea-level fluctuations (see figure). This conclusion was based on the use of the



Number of sudden changes of sea level from the 1977 (stippled) and 1987 (stripes) data of Vail and colleagues.

data of Vail *et al.* at which deliberately excluded a detailed curve of Cretaceous sea-level oscillations for proprietary reasons.

When one uses the new sea-level curve of Haq *et al.*⁵, which includes the previously unreleased Cretaceous data, the number of rapid changes of sea level in the Cretaceous is similar to that in the post-Cretaceous interval (see figure). The lack of magnetic reversals during the Cretaceous 'quiet' interval, therefore, is not matched by any lack of short-term changes in sea level during that time. Thus, the data seem not to support the Morris and Muller hypothesis. One way to resurrect the hypothesis, however, would be to suggest that during the quiet interval, the core of the Earth was less susceptible to the disturbances caused by rapid sea-level changes resulting from extraterrestrial impacts.

M. R. RAMPINO

Department of Applied Science,
New York University,
New York, New York 10003, USA

1. Muller, R.A. & Morris D.E. *Nature* **332**, 211 (1988).
2. Gaffin, S. *Nature* **329**, 816–819 (1987).
3. Vail, P.R. *et al.* *Am. Ass. Petrol. Geol. Mem.* **26**, 83 (1977).
4. Muller, R.A. & Morris, D.E. *Geophys. Res. Lett.* **13**, 1177–1180 (1987).
5. Haq, B.U. *et al.* *Science* **235**, 1156–1167 (1987).
6. Rampino, M.R. *Phys. Today* **41**, 121–122 (1988).

Accuracy of Buffon's 200-year-old experimental data

SIR—I have analysed the experimental data of Buffon^{1,2} concerning the properties of wood (breaking load, drying and dampening) and the cooling time of iron spheres extracted from a furnace and allowed to cool in air. The quality of the graphs that can be drawn from these data indicates Buffon's rigour in applying experimental method.

Georges Louis Leclerc, count of Buffon (1707–1788) conducted numerous experiments on wood^{3,4}, especially oak, particularly important at that time for the construction of ships, houses, bridges, mills and so on. Buffon wanted to verify Galileo's law on the breaking load of a plank, $P = k a b^2/L$ where P is the breaking load, a is the width, b is the thickness, L is the length and k is a constant of proportionality. He found that Galileo's law is of limited value applying better to solid, inflexible materials which reach their breaking load without previously bending. Moreover the law has greater validity when the planks are shorter. Figure 1 shows that Buffon's data for oak planks which bend before breaking, with the exception of high L values at which strong deviations from linearity occur, are consistent with the equation of the type $P = k a b^2/L^{5/4}$.

Some of Buffon's experiments on the drying of wood and its retention of humidity took more than 20 years to perform. Figure 2 shows a semilogarithmic plot of the weight of an oak parallelepiped, exposed to air but protected from the rain, as a function of time. It is possible to see some fluctuations resulting from the partial retention of humidity in winter. If Buffon had been able to compose graphs of this type he could have deduced that

water is lost from the surface and through three types of aperture in wood. Water absorbed by the surface is lost in 4–5 days (behaviour I, Fig. 2a). Water is lost from holes between the fibres in one month (behaviour II, Fig. 2a). Water from long, thin tubes running longitudinally along the trunk of the tree is lost in about eight months (behaviour III, Fig. 2b). Finally, water loss through single cells needs eight years to reach an equilibrium (behaviour IV, Fig. 2b).

In 1749, Buffon began experiments to determine the age of Earth^{3,4}, measuring

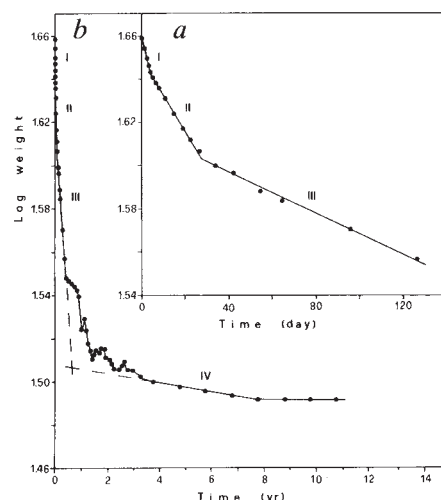


Fig. 2 Drying of an oak parallelepiped (14.2 × 8.2 × 9.4 inches).

cooling times of iron balls of various diameters, from red heat to temperatures of about 60 and 15 °C, respectively (Fig. 3). Buffon calculated that the Earth's age would be 100,696 years if it was composed solely of iron. In fact, he believed the

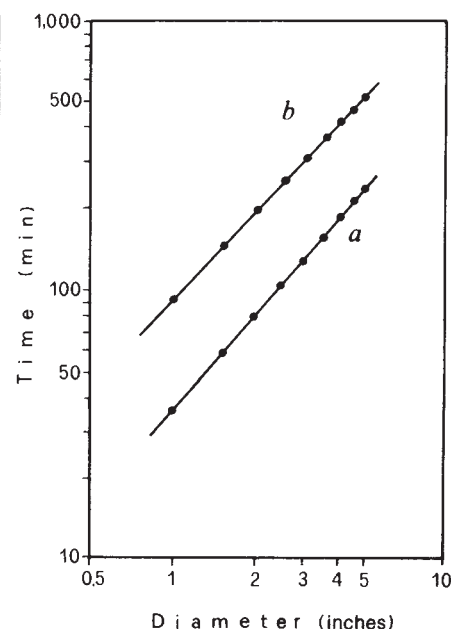


Fig. 3 Logarithm of the cooling time of iron balls with respect to the logarithm of their diameter. *a*, Cooling from red heat to about 60 °C; *b*, cooling from red heat to about 15 °C.

Earth is a mixture of glass, quartz, marble and ferrous minerals, and therefore that it should cool more quickly than iron. He then calculated the Earth's age to be 74,047 years. This is about 13 times greater than 5,750 years, the age based on calculations made in the seventeenth century by archbishop Ussher and generally accepted in Buffon's day. By assuming an average diameter of 12,470 km for the Earth, and from the slope of the straight line *b* of Fig. 3, a cooling time of 100,000 years can be calculated. This agrees completely with Buffon's calculations from the premise of an iron-based Earth.

The work I have described here confirms by the beauty of the graphs which can be drawn from Buffon's data, the rigour of this scientist in applying experimental method.

FERNANDO ZOCCHI

Istituto di Metodologie Avanzate

Inorganiche,

Consiglio Nazionale delle Ricerche,

CP 10, 00016 Monterotondo Scalo,

Roma, Italy

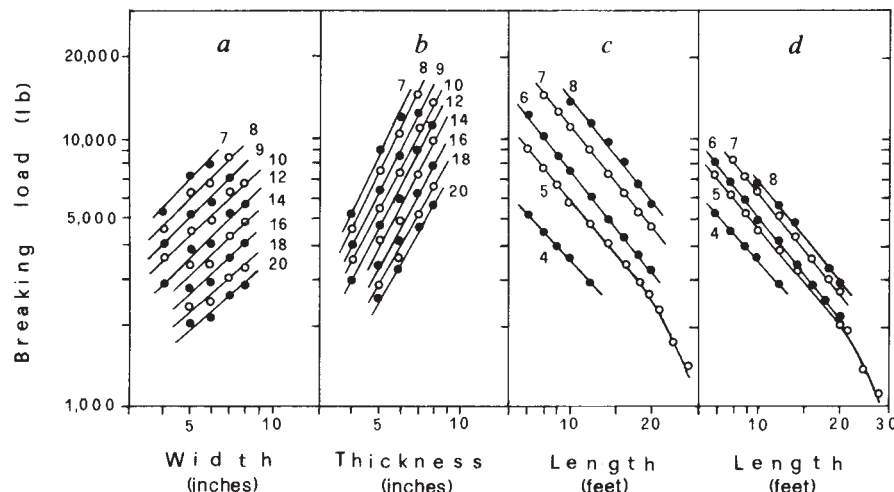


Fig. 1 Buffon's experimental data obtained from oak planks with square cross section are treated to separate the effects of the parameters *a*, *b*, *L* to clarify the dependence of breaking load: *a* on width at various lengths and the same thickness (4 inches); *b* on thickness at various lengths and the same width (4 inches); *c* on length at various thickness and the same width (4 inches); *d* on length at various widths and the same thickness (4 inches). The numbers near the straight lines indicate: in *a* and *b* the length of the planks; in *c* the thickness; in *d* the width.

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

1. Boschi, G. *La vita e i tempi di Buffon* (Napoli, 1879).
2. Piveteau, J. *Oeuvres philosophiques de Buffon* (Presses Universitaires de France, Paris, 1954).
3. Buffon *Histoire naturelle générale et particulière* (Imprimerie Royale, Paris, 1772–92).
4. Barr (trans) *Buffon's Natural History General and Particular* (London, 1792–93).

Twilight of the pencil?

William H. Press

Mathematica. Wolfram Research Inc., PO Box 6059, Champaign, Illinois 61821. \$495 (Macintosh Plus/SE version); \$795 (Macintosh II version). Stop Press: on 15 December a version becomes available for 386-based IBM-PCs; prices from \$695 to \$1,295.

Mathematica: A System of Doing Mathematics by Computer. By Stephen Wolfram. Addison-Wesley: 1988. Pp. 749. Hbk \$40.95; £36.85; pbk \$27.75, £19.95.

Mathematica is a remarkable, perhaps even revolutionary, computer program written by a small group of young programmers at Wolfram Research Inc. Inside it are of the order of 200,000 lines of source code. More interesting is that behind the program there are a number of powerful, perhaps radical, ideas on what amounts to 'machine-assisted human thought' as it applies to mathematics. The manual, available in book form, provides an intellectual entrée to these ideas; the software is separately available from Wolfram Research (for Macintosh machines) and from Sun Microsystems Inc. (for Sun workstations). *Mathematica* is included free with the new Next computers.

The program is sufficiently different from existing computer tools, and so powerful in its way, that it is hard to describe exactly what it is. The easiest description, though it misses the mark, is to say that it is a symbolic manipulation package comparable, say, to *Macysma* (Symbolics Inc.). If you type 'Integrate $[1/(1+x^6)]$ ', *Mathematica* echoes back the indefinite integral, a complicated expression involving arctangents and logarithms. As compared with *Macysma* (available for VAX, Sun and Apollo workstations, but not for the Macintosh), *Mathematica* is less good at doing integrals, but seems better at solving systems of (even nonlinear) equations, number theoretical calculations and special functions. *Mathematica* has the vastly superior user interface, but its output format (like *Macysma*'s) leaves something to be desired.

The reason that this comparison misses the mark is that *Mathematica* is deeper in its concept than previous symbolic mathematics programs. It is a kind of 'general algorithmic engine', for which symbolic mathematics is only one possible application. The engine can operate with equal facility on algebraic objects, numerical objects (arbitrary precision in integer or floating), graphical objects (represented ultimately in *Postscript* graphics language), or any other object that the user simply defines 'on the fly'. The interface among these modes is

practically seamless.

If one thinks of *Mathematica* as a computer language — another approximate but incomplete description — then it is something like LISP, but with a friendly user interface, with several dozen pre-defined data structures, and with several hundred pre-defined operations on these structures. Also pre-defined are anyone's

A nice feature is the way that your session is recorded in an on-screen 'Note-book', in which the program statements can easily be cleaned up, explanatory text inserted and the result printed out in laser-printer fonts as an elegant-looking document. Wolfram's informal statistics suggest that a surprisingly large number of users are financial analysts, who are probably using *Mathematica* in this mode.

This is flashy stuff, but by itself it is unlikely to revolutionize the way that working scientists do mathematics. Two deeper aspects of the program — which require a correspondingly greater investment of effort to master — might, however, bring about just such a change.

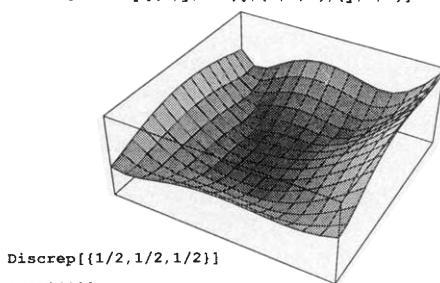
The first of these is *Mathematica*'s usefulness as a concise test-bed for algorithmic experimentation. Some sense of this may be conveyed in the accompanying figure,

even without detailed explanation. A sequence of definitions (lines with the ':=' construction) creates eight 3-vectors on the corner of a unit cube, defines 'triangular' linear interpolation functions, and teaches *Mathematica* to compare the inverse square force resulting from a particle at an arbitrary location with that from the particle's interpolation onto the corners of the cube. The definitions are then used in three ways: first, to make plot of the force comparison on a two-dimensional slice through the unit cube; second, to compute a particular value; and third, to give an algebraic formula. It would be difficult, if not impossible, to get the same intuitive grasp of a problem like this by hand calculation.

Notice that nothing in the definitions specifies whether they are numerical or symbolic. That is determined, interpretatively, by the actual calling arguments in each invocation. This, in fact, relates to the second deep aspect of *Mathematica*. It has a sophisticated hierarchical mechanism for 'overloading' procedural definitions, even its most fundamental pre-defined ones. The same operation (addition, for example, or 'do loop') can be defined to work on numbers, vectors or any other object, in any desired way. Functions automatically parse their arguments and search for a previously stored definition that matches syntactically. If nothing matches, then the function becomes a new symbolic 'atom'. Once one masters the knack, one can 'teach' *Mathematica* to be anything one wants it to be, or to deal with virtually any kind of mathematical object.

The potential user should be aware of some practicalities. *Mathematica* requires at least four megabytes of memory to run effectively on a Macintosh — and Mac

```
Do[ p[4i+2j+k+1] = {i,j,k}, {i,0,1},{j,0,1},{k,0,1}]
Norm[p_] := Sqrt[p[[1]]^2+p[[2]]^2+p[[3]]^2]
w[i][q] := Product[1-Abs[q[[j]]-p[[j]]], {j,3}]
Force[p1,p2] := (p1-p2)/Norm[p1-p2]^3
GridForce[q] := Sum[ w[i][q] Force[p[i],ReceivePt], {i,8}]
ReceivePt = {2.6,1.9,1.0};
Discrep[q] := Block[ {F1,F2},
  F1 = Force[q,ReceivePt];
  F2 = GridForce[q];
  Norm[F1-F2]/Norm[F1]
]
Plot3D[Discrep[{x,y,0.5}],{x,0,1},{y,0,1}]
```



```
Discrep[{1/2,1/2,1/2}]
0.0106823
ReceivePt = {d,1/2,1/2};
Discrep[{1/2,1/2,1/2}]/InputForm
(1/2 - d)^2*((1/2 - d)^(-2) + d/(2*(1/2 + d^2)^(3/2))) -
(1 - d)/(2*(1/2 + (1 - d)^2)^(3/2)))
```

Using Mathematica for algorithmic experimentation.

favourite control structures, an inclusive set taken from C, FORTRAN and Pascal. *Mathematica* emerges as a higher-level language than any of its antecedents. It is a huge language, in terms of its number of pre-defined operations.

All of this structure means that *Mathematica* can be used by the novice as a kind of super-BASIC. Follow the examples in the book, and you have: (i) a numerical calculator that knows lots of special functions, handles real or complex numbers in arbitrary precision; or (ii) a graphics package that plots functions, contour plots, shaded three-dimensional surfaces (in colour if you like); or (iii) a symbolic mathematics package that can do algebra, matrix algebra, derivatives, integrals and so forth.

memory is very scarce at the moment. Its speed on a Sun or Macintosh II is more than adequate, but the version for Macintosh Plus or SE is likely to be frustratingly slow. As with any new software *Mathematica* is not without bugs, so upgrades are essential; these are free for 90 days, but will require a (modest) continuing investment thereafter.

These caveats aside, *Mathematica* is a startlingly good tool. Research groups

with Suns or Macintosh IIs should acquire it. Indeed, the program is, by itself, nearly enough justification to buy the necessary hardware. There is no reason in principle that it cannot also be made available for VAX workstations and 386-based MS-DOS (but not 8086 or 286) machines; I hope that happens soon. □

William H. Press is at the Harvard College Observatory, Harvard University, Cambridge, Massachusetts 02138, USA.

Moving story

Roger Lemon

The Neural and Behavioural Organization of Goal-Directed Movements. By Marc Jeannerod. Clarendon: 1988. Pp.283. £32.50, \$49.95.

C. S. SHERRINGTON wrote that "To move things is all mankind can do; for such the sole executant is muscle, whether in whispering. . . or in felling a forest". Understanding the organization of movements remains a central issue in research on brain and behaviour. And, by common consent, attempts to do so require theoretical and experimental studies of a truly multidisciplinary character.

Marc Jeannerod's book is an excellent example of what can be achieved in this way. Its publication marks a point when the many approaches to the problem of understanding movement are rapidly converging. Among them is the use of inverse kinematics to predict the way in which a movement is initiated and controlled by studying its dynamic features. Jeannerod also draws on experimental evidence from two other fields: the neurophysiology of behaving animals, in which neuronal activity in motor areas of the brain is monitored during movement; and the use of cognitive psychology to study motor control in human subjects. Jeannerod examines how these approaches have furthered our knowledge of the organization of goal-directed movements. A good example is the visual guidance of the hand towards a specified target. Jeannerod has made notable contributions in this area, and the best parts of his book are those in which he uses this type of movement to illustrate some basic principles of motor organization.

The book opens with a consideration of the main theoretical arguments for a central representation of motor 'programs'. As Jeannerod writes, "Movement is not a transparent phenomenon", thereby neatly stating a central problem for those striving to understand movement and its control by the brain. When compared with the working of our sensorium, we have relatively little insight or perception of how we produce and control

movements. Ideas about how the brain organizes movement have a long and interesting background, well discussed in Jeannerod's earlier book, *The Brain Machine*. Here he shows how hierarchical theories of neural organization have provided the main theoretical framework for investigation of the motor system. In the 'top-down' model of motor representation, Marr proposed a hierarchy consisting of a highest level of 'computational theory' which defines the goal of the computation; an intermediary level which transforms the computed input to an output; and finally a third, hardware level for implementing the output. But we now know that even the most elementary motor processes can be influenced by specific cognitive states or behavioural 'set', and this makes the identification and separation of the different hierarchical levels extremely difficult.

Jeannerod is an excellent guide to this difficult area. His account is readable and full of direct experimental evidence for a hierarchical organization. In visuo-motor reaching, for instance, there is evidence that the central command for moving the eye, head and hand occurs simultaneously; the resulting movements have successively later onsets largely because the mechanical properties of these different segments are so different. Earlier authors have not shrunk from including, without much justification, the term 'control' in the titles of their books. Here, Jeannerod explicitly attacks the question of what it is that is being controlled, his examples including the direction and amplitude of movements and the co-ordination of the proximal (arm) and distal (hand) components during reaching. This co-ordination, although very precise, is controlled by different neural mechanisms.

Although knowledge of the neural machinery that underpins movement has resulted mainly from animal studies, our understanding of how this neural machinery is put to work before and during voluntary movement has come increasingly from experiments in human beings. These studies have combined ingenious design with new techniques for registering movement. Jeannerod's approach has been radically different to that of much of the work on animals, where there has been a tendency to

"reduce complex movements to units functioning in the same mode as single joint movements". In a strong plea for more research into multi-segmental actions, such as eye-hand control, he points out that "parameters currently considered in the study of single-joint movements, such as latency, duration, error with respect to the target and the kinematics of the end point of the limb are far from representing an adequate description of motor performance". Investigations of complex movements show that their trajectory is somehow abstracted from the movements at individual joints, and these studies have revitalized the quest for the neural mechanisms which can perform such an abstraction. In this regard it is remarkable that there is no mention in the book of the work of Georgopoulos and his colleagues on the activity of monkey cortical neurons during reaching.

In some of the most controversial areas Jeannerod is refreshingly non-partisan, preferring to present the available evidence in a logical sequence. The theoretical territory of many areas of motor control has all too often been staked out by the protagonists of mutually exclusive theories, and experimental results have been manipulated to fit one or other of these theories. It has long been argued, for example, that feedback control of movement depends either on the afferent information arising in peripheral receptors, or on internal feedback/corollary discharge mechanisms, by which the brain is constantly informed of movement by monitoring its own command signals. But sensory and internal feedback mechanisms cannot be considered to be mutually exclusive, because the former provides the rules by which the latter operates once a particular movement has been learned. As an example, Jeannerod reviews the evidence showing that disruption of sensory feedback from the eye muscles produces errors in reaching for visible targets. On a separate line of reasoning, he shows that the maps of proprioceptive and visual modalities must remain in register with one another in order for them to interact for successful reaching into extrapersonal space. In the last chapter he discusses the pivotal role of the posterior parietal cortex in co-ordinating poly-sensory inputs for the control of reaching.

Although this book will be of particular interest to those studying visuo-motor behaviour, it deserves to reach a much wider audience. It has a clarity of style and of argument that makes it a rewarding read for all those interested in understanding how the brain organizes such tasks as forest-felling and whispering. □

Roger Lemon is in the Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK.

Muller's period piece

Antoni Hoffman

Nemesis: The Death Star. By Richard Muller. Weidenfeld & Nicolson, New York:1988. Pp.193. \$17.95. To be published in Britain next year by William Heinemann.

ON ITS jacket this book bears the legend *The Story of a Scientific Revolution*. The revolution Richard Muller aims to describe refers, of course, to the idea that periodic catastrophes are brought upon life on Earth by comet showers thrown towards the inner planets by an unseen solar companion, Nemesis. And the readers are told that here they will find *the*, rather than *a*, story of this idea because Muller is the inventor (though not yet discoverer) of Nemesis, and one of the collaborators of Luis Alvarez who proposed the impact hypothesis to explain the mass extinction at the end of the Cretaceous — the hypothesis that made it possible (though by no means easy) to think of Nemesis.

The development of the Nemesis hypothesis is indeed a fascinating intellectual adventure, clearly showing the power of trespassing across the established borders between scientific disciplines. Undoubtedly, a geologist or palaeontologist might be able to break with his subject's tradition and explain mass extinctions by periodic comet showers, but he would not be able to develop a plausible astronomical theory; yet, without such a theory, be it one involving Nemesis or Planet X or perhaps something else yet to be proposed, the explanation would not even be remotely acceptable to specialists. But how does a successful physicist become interested in problems of the history of life on Earth? How does he get around his lack of expertise and even basic understanding of geology and palaeontology? How does a new idea originate, become published and publicized, develop into a research programme? It is indeed exciting to have a first-hand account of the process.

As a storyteller, Muller is excellent. His book vividly describes the ways in which science is done. It portrays all of the main players in the game of the end-Cretaceous impact and Nemesis hypotheses as real characters, for whom Muller has sympathy or dislike. It shows how many false and even ridiculous solutions to a problem are considered and evaluated before the plausible ones are hit upon. It points to the importance of maintaining both scepticism and enthusiasm while working on the frontiers of research. And it emphasizes the crucial role of informal contacts, ties of friendship, mutual confidence, and even

gossip and hearsay in the progress of science — for this is the context in which new ideas originate and, well before they are formally published, undergo evaluation and modification, stimulate further studies and generate research funds. When summarized in such general terms, this sounds nothing special. The interest is in the detail, however, as in the story of Muller learning from David Raup what rival astrophysical articles on periodic mass extinctions Raup is receiving as a referee for *Nature*.

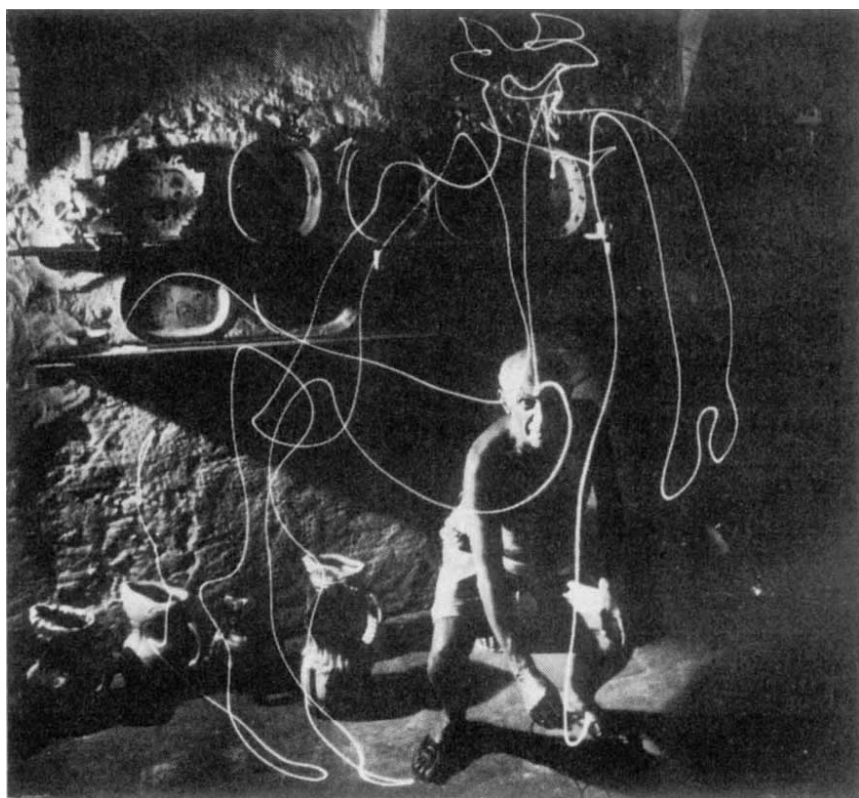
On top of all this, Muller's book is a very good read. Perhaps even too good, for it is easy to miss its clear intention to convince rather than to describe. Never mind the many minor, though annoying, mistakes (for example, Thomas Kuhn is an eminent physicist and philosopher not a 'writer'; coccoliths are remains of algae not animals; and the textbook *Principles of Paleontology* is co-written by Raup and Steven Stanley, who thus share the responsibility for the assertion that "the causes of mass extinction are not simple"). More importantly, Muller presents an extremely one-sided argument and his description of his opponents sometimes descends to caricature. In his view, William Clemens, whom he condescendingly calls "the local dinosaur expert", just "dug in his heels" when proved wrong on the geology of dinosaur-bearing strata in Montana. Muller does not entertain the

possibility that the alleged proof might be wrong, as may in fact be the case, depending on one's interpretation of subsequent sedimentological studies.

Similarly, Muller tells how Charles Officer claimed that mineral microspherules are not restricted to the Cretaceous-Tertiary boundary clay in Italy, and how this claim was ridiculed by Walter Alvarez who found modern insect eggs posing as microspherules at rock surfaces of the section concerned. It does not even occur to Muller that alongside Recent contaminants other microspherules may also abound in the section, as has indeed been demonstrated in Italy as well as in Spain and Denmark. Muller notes that the Nemesis hypothesis depends on the postulated periodicity of extinction, but he fails to mention the counter-arguments to the concept of periodicity, as if there were no statistical problems whatsoever.

Nemesis is certainly fun to read. But those who would like to examine the facts and arrive at their own judgement of events should instead turn to David Raup's *The Nemesis Affair*. Raup's book is written for the same broad readership and from the same partisan position, but it offers a more balanced view and more expert treatment of the pros and cons for Nemesis. □

Antoni Hoffman is at the Institute of Paleobiology, Polish Academy of Sciences, 02-089 Warszawa, Poland.



Light work — Picasso's creation of an image of a bull, captured with time-exposure photography. This example of modern abstract symbolism is compared with examples reaching back 300,000 years in Roger Lewin's *In the Age of Mankind*. The book is a large-format, illustrated account of the age and complexity of human ancestry, and takes in the background and significance of many recent discoveries in anthropology. Publisher is Smithsonian Institution Press, price is \$35.

Something in the air

S.A. Penkett

Chemistry of the Natural Atmosphere. By Peter Warneck. *Academic*: 1988. Pp.757. \$85, £63.

ALTHOUGH scientific observations of the composition of the atmosphere were made in the 1700s and earlier, the real roots of atmospheric chemistry lie in the studies of ozone and the acidity of rain-water carried out in the last century. The subject has matured only slowly but today atmospheric chemistry stands as a recognizably distinct discipline, not the least of the indicators being the increasing amount of attention publishers have paid to it.

Two books published in the early 1960s set the pattern for later authors. One, *Photochemistry of Air Pollution* by Philip Leighton, dealt with the subject from the point of view of a laboratory photochemist. The other, Christian Junge's *Air Chemistry and Radioactivity*, concentrated on presenting and interpreting observations of the meteorological and chemical behaviour of trace gases and aerosols in a global atmosphere. Peter Warneck's book fits into this second mould — not surprisingly perhaps, because Warneck was for many years a co-worker of Junge at the Max Planck Institute for Chemistry in Mainz.

Much has changed since the first appearance of Leighton's and Junge's books. For instance the original hypothesis proposed to account for the chemistry of stratospheric ozone has been displaced by theories emphasizing the role of minor trace gases, including nitrogen oxides and halogen compounds, in catalytic destruction cycles. Also in the troposphere, a universal oxidation mechanism has been discovered whereby most trace gases emitted from the Earth's surface are destroyed by reaction with free radicals and peroxides before they can reach the stratosphere. Both in the stratosphere and the troposphere, therefore, atmospheric chemists are mainly concerned with the role of trace gases, with their emissions, their distribution and their chemical interactions, which in many cases are initiated by the absorption of sunlight.

Warneck's text deals predominantly with such phenomena as they occur in the troposphere, and is more thorough than many others in treating both gaseous and particulate species in similar depth. It also covers the atmospheric aqueous phase extensively, both from a physical and a chemical point of view. There is much fascinating detail on observational data which not only makes the book unusually interesting to browse in, but, with its

numerous and useful tables, a solid work of reference.

Perhaps logically, given its title, *Chemistry of the Natural Atmosphere* does not deal primarily with many forms of air pollution, although there is a section on photochemical air pollution in the chapter on ozone in the troposphere. The ozone hole was found too late for it to be mentioned in the chapter on stratospheric chemistry, and the stress in the chapter on carbon dioxide is much more on geochemistry than on the radiative problems associated with its increasing abundance. Chemical cycling through the atmosphere determines the structure of other chapters on individual compound classes, and there is a nice discussion on the evolution of the atmosphere to conclude.

One area that is not well covered is the influence of droplet and particulate

phases on the homogeneous gas-phase chemistry, a topic that is becoming increasingly relevant both for the case of cloud droplets in the troposphere and for ice particles in the stratosphere which increase the efficiency of ozone removal over Antarctica. The omission is partly, but not entirely, explained by the fact that the literature survey — which is otherwise reasonably thorough — largely ceased in 1985.

This is a book which has several unique features, and which will be of interest to most atmospheric scientists. In particular, the material is ideal for teaching post-graduate students about the chemistry of the natural atmosphere and its response to increasing emissions of pollutants. □

S.A. Penkett is in the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

Solid solutions

Paul W. Williams

Geomorphology and Hydrology of Karst Terrains. By William B. White. *Oxford University Press*: 1988. Pp.464. £35, \$45.

WILLIAM WHITE adopts the Soviet view that landforms produced by solution processes constitute karst. By that definition about 20 per cent of the Earth's land surface can be considered karstic, and this book deals with the landforms and hydrology of such terrains.

Until the 1960s karst research was largely the preserve of Europeans. North Americans made comparatively few contributions to the subject, despite the fact that about 20 per cent of the coterminous United States is composed of carbonate rocks, with additional huge areas of even more soluble evaporites. Since then the situation has changed markedly, and scientists such as William White have risen to international prominence.

White and his associates are particularly well known for their work on carbonate geochemistry and groundwater hydrology, and especially for linking solution kinetics of limestones to cave and aquifer development. From a base in Pennsylvania State University, the extensive karsts of the nearby Appalachians and Kentucky have provided numerous field sites to test and develop hypotheses. This volume largely records that experience, which has accumulated over more than 25 years.

The form of the book is closer to a series of essays than to an account integrated by a pervading theme or conceptual approach. Consequently, it will be of greater value for professionals in related fields than for undergraduates. This is in keeping with one of the author's primary

objectives, that of writing a book suitable for water-supply specialists, urban planners and environmental engineers. The choice of examples also makes it best suited for a North American readership.

White identifies three essential features of the karstification process: the hydrogeological setting, the chemical driving force and the physical driving force. In developing these points, he provides notable accounts of carbonate dissolution, the mechanics of groundwater flow and the geochemistry of karst waters. His command of these fields on occasion results in authoritative pronouncements on the linkages between laboratory theory and field evidence. This is a rare talent and is well exemplified here in the discussion of the processes of karst conduit and aquifer development. White is less successful in forging a link between chemical-hydrological theory and surface landform development. So the geomorphological content is comparatively conventional, except in the section on karst and karst-like features in slightly soluble rocks.

The concluding chapters concentrate on the application of our knowledge of karst to a range of practical problems. Water resource contamination, foundation engineering, sinkhole collapse and flooding are all the subjects of broad and useful surveys that will be especially helpful to environmental managers. □

Paul W. Williams is in the Department of Geography, University of Auckland, Private Bag, Auckland, New Zealand.

New in paperback

- *Thermoluminescence of Solids* by S.W.S. McKeever. Publisher is Cambridge University Press, price is £17.50, \$34.50. For review see *Nature* 317, 122 (1985).
- *Dinosaurs Past and Present*, Vols I and II, edited by S.J. Czerkas and E.C. Olson. Publisher is University of Washington Press, price is \$22.95 per volume. Available in Britain from British Museum (Natural History), £19.95 each.

Investment with an eye to the future

Switzerland's prosperity, based on intelligent exploitation of technology and business acumen rather than natural resources, leans heavily on success in research and education.

SWITZERLAND is a narrow strip of the foothills between Northern Europe and the Alps, with a poorly connected maze of mountain valleys to the south. The native population is roughly 6.5 million people (half as many again as in Ireland), most of whom seem to speak at least two out of four languages (French, German, Italian and Romansh, not to mention English).

To keep the wheels of prosperous industry turning, there are also at least 1 million foreigners, many of them technically qualified people, but this number does not include those who work for the international organizations that have sought out Switzerland as the most probable permanently neutral country.

The most striking feature of Switzerland is that it is truly tiny. (Ireland seems larger because the roads are not as good.) Hardly any city (of which there are five with more than 250,000 people) is more than two hours driving from Berne (in French) or Bern (in German), the administrative centre that would elsewhere be called the capital city. But sensible people take trains, which connect the same cities frequently, quickly and more cheaply and also serve a high proportion of the villages. (So well ordered is the place that, if there is no train, there is almost certain to be a connecting bus from the nearest station.)

The next most striking feature of the country is how rich it is. In terms of gross domestic product on a per capita basis, Switzerland ranks with the United States and Sweden. But the economy seems consistently to find a healthy balance. The inflation rate is meagre (2.21 per cent on the average over the past six years), the Swiss franc (SFr1.00 = \$0.69 at present) is a hallmark of stability, the external balance of payments is consistently positive and in 1987 people saved 8.4 per cent of their incomes. But the simplest way of telling that the country is prosperous is to

notice how quickly one's spending-money melts away, which is a way of saying that services are relatively costly (which is a way of redistributing income in the absence of taxation-based welfare).

Yet the real marvel of Switzerland is notoriously its constitution, by which 26 autonomous cantons (the last, Jura, joined only in 1979) delegate to the confederation a limited range of powers — defence primarily, foreign relations only with the proviso that there should be hardly any (which is why Switzerland does not belong to the United Nations), the administration of justice, the determination of citizenship and the continuing process of interpreting and amending the constitution.

The confederation in Berne is largely financed by a levy on the taxes cantons themselves raise. It is managed by a council of seven people elected in rotation by the two elected houses of the parliament in Berne. Most government functions are conducted by commissions reporting equally to the federal council and the parliament. The confederation, neither council nor parliament but their amalgam, can win extra powers only by putting a draft-enabling law to a countrywide referendum (in which women can have their say even in cantons such as Appenzell where they have no vote in local elections).

The pattern of Swiss science mirrors the constitution. Higher education, for example, is strictly the responsibility of the cantons, only eight of which have universities (Basle, Berne, Fribourg, Geneva, Lausanne, Neuchâtel, Zurich and the business school at St Gall). The confederation maintains two polytechnics (at Lausanne and Zurich). Italian-speaking Switzerland has no university. The confederation has slowly extended its right to sponsor research at the cantonal universities, first by means of grants from organizations such as the Swiss National Science Foundation (see page 326), more recently by direct subvention as well. Such national research centres as there are have grown up by accident beneath the wings of universities or have required draft laws and referenda.

The natural suspicion that, in such circumstances, Swiss science would be hamstrung by lawyers and bureaucrats appears, however, to be unfounded. The simplest way to avoid trouble with the lawyers is to obey the law (which, in Switzerland, is not so aloof as to allow

academics to strike deals with their universities about the dates at which they retire from work, for example). Inured to compromise, people seem not to ask for the Moon (or for extravagant research grants) and then complain when they do not get their way. And the bureaucrats pride themselves on having absorbed the Swiss business tradition that a bargain can be sealed by word of mouth.

So why should Switzerland just now be as alarmed about the future as it seems to be? The collapse of the Swiss watch industry in the 1970s, in the face of competition from Japan and the Far East, has clouded people's confidence that all will be well. Bureaucrats and academics also recall that, at about the same time, the confederation could make little progress with its attempts to modernize the infrastructure of research and higher education: referenda repeatedly rejected enlightened legislation. Now, looking back, people recognize that Switzerland, which might have been Europe's software house by now, was painfully slow off the mark in computer technology. How could a literate and technically advanced community used to running cottage industries and multinationals have been so blind to important and foreseeable innovation? Will the waywardness of the cantons and the voters prevent desirable change at other crucial times in the future?

The European Community's determination to become a truly single market in 1992 is the next big threat. Will Switzerland become an isolated island in an industrial market with 50 times as many people? There is some loose talk of applying to join the Community, but that is plainly impossible: the constitution would be the first casualty of membership, for the people at Berne do not at present have the power even to negotiate with Brussels about the weight of the trucks allowed to cross Switzerland by road (one current lively issue). Membership of the Community would also require Switzerland to allow anybody from elsewhere in Europe to live there, which is unthinkable in the present climate of opinion. That no doubt explains why Swiss industry is said to be marching with its feet, buying up companies inside the boundaries of the single market, which could be bad news of another kind.

A more insidious danger is the shortage of skilled people now emerging. The one million workers from abroad are a reproach, but applications for engineering courses at the polytechnics remain less

Science in Switzerland

THIS brief account of scientific institutions and developments in Switzerland has been written by Steven Dickman (Munich correspondent) and John Maddox. The authors acknowledge that they have been able only to scratch the surface of a vigorous research community. They are grateful to the many public servants, academics and research scientists who have helped them with information. □

buoyant than applications for courses in law and accountancy. The planners seem to be seized of the dangers, but it remains to be seen how successful they will be in their efforts to recruit young people into science and technology and to challenge Swiss manufacturing industry with innovation. The dilemma is that the financial institutions which seem to be the magnets for young people are also potential sheet-anchors for the years ahead.

That, the optimists say, is why the future need not be scary. The banks can ensure that Swiss industry has the capital to remain competitive, in which case it will always be possible to recruit engineers from elsewhere, usually in Europe. There are also the successful Swiss multinationals, which accomplish the same ends by manufacturing what they sell elsewhere. But that is where the example of the watch industry seems threatening. What would happen if that upset should be repeated, perhaps on a larger scale?

This brief survey of science in an exceedingly interesting small country has mostly nothing but admiration to report. The Swiss seem mostly to be doing sensible things. Among other things, they spend a greater proportion (2.9 per cent) of their gross domestic product on research and development than most other countries except Japan, the United States and Sweden. But there is one respect in which Swiss policy on science and higher education seems dangerously complacent: the proportion of the young going on to higher education courses is smaller (at 12 per cent of the total, with 15 per cent of men and only 8 per cent of women) than in comparable countries such as the United States and Japan.

The Swiss are quick to say that the comparison is misleading because the cantonal technical colleges continue to follow their traditional role of preparing technicians and engineers for employment (but women are also under-represented among recruits to those courses). Everybody agrees that this system has served Switzerland well. But the counter-argument is not especially forceful. If future industry is fashioned from technology not now foreseen, how can a traditional system of technical training substitute for higher education informed by adventurous research? Switzerland is not the only modern state in which this point could usefully be taken, but it is one of the few with resources ample enough to meet whatever demands may emerge, and with the will to do so. The snag is simply that there is no way in which the Berne confederation can decisively influence the decisions of young people about their careers. In the end, it will probably fall to Swiss industry, which already pays for 80 per cent of Swiss research and development (the highest proportion anywhere) to take up the task of persuasion. □

Political structure

A well-ordered system

SWITZERLAND may be geographically tiny, but even by those standards, the Swiss confederation is tinier. The government's annual revenue is something like 10 per cent of the gross domestic product, compared with about 40 per cent in Britain and 20 per cent in the United States. That the government can afford to spend anything on science is, in the circumstances, surprising and creditable. But, in reality, it falls to private industry to carry most of the burden, close on 80 per cent of the total.

Out of an estimated spending on research and development of SFr7,000 million in 1986, Swiss industry is reckoned to have spent 78 per cent, most of this in its own laboratories. By contrast, the confederation is reckoned to have spent 16 per cent, most of that in higher education institutions and their associated laboratories. The cantons are credited with some 6 per cent of the total, but that consists mainly of spending on researchers' salaries and their equipment.

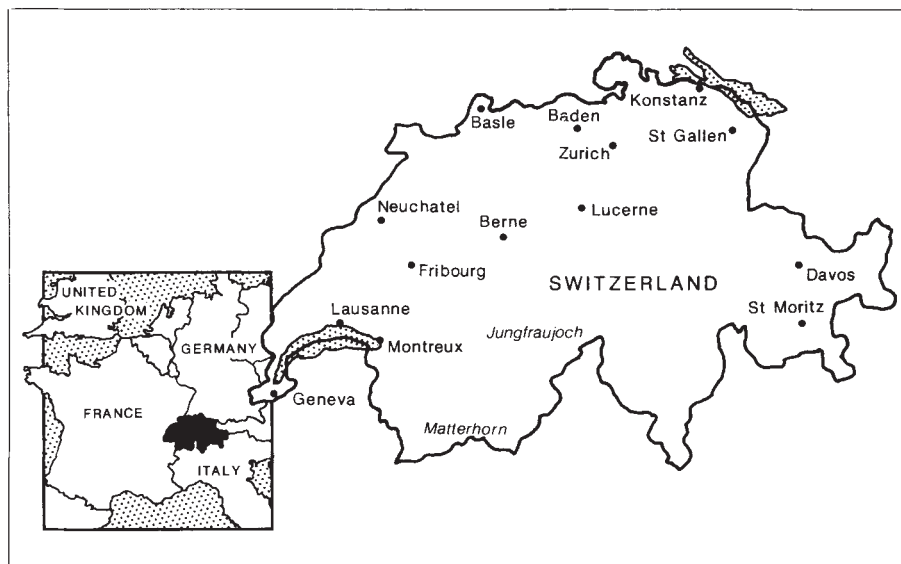
Notably, this pattern most of all resembles that in Japan, where industrial spending on research and development is much greater than that of the central government. Government officials remark uneasily, however, that during the present decade, there has been a tendency for major companies (especially in the chemical industry) to increase their research and development expenditure outside Switzerland, and at the same time to reduce the value of their contracts with universities and other institutions of higher education.

The same officials are encouraged by the relatively generous spending on research and development by smaller corporations. Although the totals are small (less than SFr200 million a year), corpora-

tions with fewer than 50 employees seem to have substantially increased their spending on research and development as a proportion of their turnover, to an average of more than 10 per cent in 1986. This seems to reflect a growth in the number and the aggressiveness of small high-technology companies.

The confederation's own spending goes in a greater variety of directions. The chief conduit for the government's spending is the Office for Education and Science, with a budget of SFr717 million in the current year. Nearly a third of that (SFr195 million) will go to the independent Swiss National Science Foundation (SNSF), chiefly a grant-making foundation whose funds are mainly directed towards academic scientists who (in Switzerland) may be social scientists (but there is a separate foundation for the humanities). For the rest, the office is also the channel for subventions of the cantonal universities and for the cost of the recently established microtechnics institute (intended to put the watch industry back on its feet) at Neuchâtel.

For the rest, individual government agencies spend small sums of money on their own affairs. Among these, the foreign affairs department has the important role of negotiating agreements with international agencies such as CERN (the European centre for high-energy physics on its doorstep at Geneva), the European Space Agency (where Switzerland has a particular interest in the space-station shuttle *Hermes*) and some of the projects of the French-inspired EUREKA high-technology programme. But the biggest spenders on research and development among the government agencies are the military services, which acknowledged research and development spending of



SFr132 million in 1986 (but of which next to nothing else is known).

Predictably, in circumstances (again like those of Japan) in which consensus is of the essence, there is an elaborate network of advisory committees. Every agency with authority to spend money on research and development appears to have one — the federal council, for example, has a 'Permanent Commission on Science and Technology', the SNSF has several (as befits a grant-making body).

There are also several layers of planning committees. By the standards of less well-ordered countries elsewhere, the Swiss process is extraordinarily deliberate. In the spring, all claimants on the federal budget have to submit a forecast of spending for the four-year period, two

years on, together with a justification.

At different levels in the structure, other committees submit their own views of the directions in which research should be encouraged, as a result of which the federal council will eventually draw up a list of priorities. Perhaps more than six months before the year begins, the parliament will vote the funds of those agencies whose bids have been successful.

In a country in which inflation is negligible, the result is that operating agencies know where they stand well ahead of the time at which they have to begin spending money. By the same test, the confederation knows at what rate to tax the budgets of the cantonal governments so as to balance its own books. What could be better ordered? **J.M.**

National Science Foundation

Is the butter spread too evenly?

GRANT-making agencies invoke a mixture of love and hate in those who depend on them, but the Swiss National Science Foundation (SNSF) seems to have won less than its expected share of the latter. Part of the explanation is the foundation's legal status, which makes it independent of the government. Part of the explanation is that its budget appears to satisfy most of the demands upon it.

But more important still is the way in which the foundation has been managed with a light touch during the past eighteen years by Dr Peter E. Fricker, distinguished as a planetary geologist (mostly during a long spell with NASA in the United States) during the 1960s.

There is a sense in which the SNSF is still marked by the apparent needs of the time at which it was established, in 1952, as the confederation's chief agent for the sponsorship of research. In those days, the creation of research centres was a part of the foundation's task. Now, that functions rests chiefly with the Office of Education and Science (see next page), on which SNSF has itself become dependent. Since 1968, it has concentrated chiefly on making research grants.

Like other grant-making foundations, SNSF uses sectional committees and external referees to decide which projects shall win its grants; as in German-speaking Europe generally, "science" includes social science, which in turns includes some history.

The budget (on a rising trend in real terms since 1981) is just under SFr200 million during the current year, but can in principle be increased by drawing on the foundation's own income, the interest on the accumulated unspent balances of previous years. Typically, the success-rate of the foundation's applicants is higher than painful experience elsewhere suggests should be the norm at about 80 per cent.

With spending on grants running last year at SFr150 million and covering 820 successful (but sometimes scaled-down) applications, the average value of the foundation's grants was just under SFr200,000. The duration of the grants awarded seems to vary enormously, from six months to more than three years.

SNSF prides itself on the informality of the application process. People are not expected to send inch-thick bundles of copied applications; two will suffice. The decision-making process is where the fun begins, in that there are not merely external referees but a group of research committees either centred on the ten universities or the scientific academies, the Swiss National Science Academy, for example. The second layer of peer-review is meant, among other things, that an applicant's proposals are feasible. There are then three sectional committees which make recommendations on particular applications to the foundation's council. Perhaps the claim that the process is completed within six months of the two annual deadlines for applications is, in reality, a boast.

SNSF has a range of other programmes — publications grants, for example — but two stand out. There is a fellowships programme costing SFr17 million a year is used to support a few hundred younger scientists within a few years of their doctorates for short periods, the majority of them at research centres outside Switzerland. (This is a curious inversion of the policy of the 1960s, when the foundation maintained a long-term fellowship programme designed to keep Swiss scientists at home.)

The foundation also spends some 15 per cent of its resources (SFr28 million last year) in support of what are called the National Research Programmes — projects of an interdisciplinary nature chosen

by the Science Council (the immediate adviser of the federal council) which range from hard ("applications of artificial intelligence") to soft ("Switzerland in a world of change") which are carried out in universities.

Broadly speaking, the foundation's customers in the universities are happy with its performance. It seems especially important that it is possible to reach officials of the SNSF by telephone, to clarify a budget item or to discuss a shorter period of grant. There is also a widespread sense that applications are dealt with fairly — not surprising given that four-fifths of the applications succeed at some level. It is also apparent, in the universities, that grants by the SNSF account for a substantial part of most institution's support from external sources. Even the two polytechnics rely to a large extent on the foundation for external research support, last year's to the tune of SFr22 million at Zurich and SFr47 million at Lausanne.

The most obvious difficulty with the smoothly working arrangements is that the funds available are too thinly spread. Would it not, many researchers say, to concentrate what funds there are on the support of a smaller number of more important and exciting projects? Part of Fricker's answer is that SNSF does just this already. Indeed, in spite of having turned its back, twenty years ago, on the creation of new research laboratories, it has used its grant-making function to foster important laboratories such as the Space Research Institute at the University of Berne, still supported with more than SFr1 million a year and known, in accordance with the confederation's plan for science, as a centre of excellence.

But that is not the general rule. The much more common pattern is that the foundation's research grants are bread-and-butter grants, essential to the well-being of research groups in all disciplines and most Swiss universities. So much can be told from the proportion of the total grant-money eventually spent on salaries, which increased from 64 per cent to 76 per cent between 1971 and 1985. (By comparison, the proportion of foundation funds spent on equipment has fallen from 17 per cent to around 7 per cent, leaving the cantons to take up the slack in the cantonal universities.)

The result is that, while a large proportion of the grants awarded at the twice-yearly adjudications are for periods not much longer than a year, recipients count on successor grants to keep their research alive. The general approval for the foundation's work suggest that researchers are not often disappointed.

This pattern is no doubt reinforced by the general understanding that SNSF support is a necessary condition for the survival of academic research programmes, at least in the cantonal univers-

ities. While the universities are generously provided, so that academics have time at their disposal and are able to spend university funds on equipment and supplies, the formality of the rules governing the appointment of people to university staffs means that SNSF is an essential source of extra help.

Part of the explanation is that the volume of the foundation's support is being steadily eroded by (albeit moderate) inflation. While nominal confederation support of the foundation increased from SFr 127 million to SFr 197 million in the decade to last year, in real terms this amounts to an increase of rather less than 20 per cent. The pressure on the foundation has been increased by the disproportionately rapid growth of academic salaries during the same period.

A further difficulty is the scope of the foundation's interest, which encompasses mathematics and engineering at one extreme and clinical medicine at the other.

This pattern of generous support spread uniformly is no doubt further reinforced by the decision-making process. While the foundation takes care that referees' opinions on grant proposals are recruited from foreigners as well as Swiss, the decisions are ultimately made by the members of the three subcommittees (called divisions) of the foundation's science council — people who, inevitably, are familiar with the applicants. Moreover, the continuing research they are asked to support, if not world-shaking, is almost always of high quality. How, in these circumstances, can one say "No"?

During the past few years, the foundation has been campaigning for a larger budget, partly on the grounds that it is now especially important to support research by middle-rank academics in preparation for the two decades ahead, when it is estimated that an unprecedented proportion of the academic community will be reaching retirement age. But, so far, its campaign has not been especially successful — this year's subvention from the confederation is slightly less than that last year.

The involvement of SNSF with what are called the National Research Programmes (overseen by a fourth subcommittee of the science council) is very different. These schemes are similar in organization to the *Sondersforschungsbereichen* of the West German research council DFG. Projects are approved by the responsible subcommittee of SNSF and by its council on proposals from a variety of sources. Project leaders are chosen and required to organize programmes of collaborative research involving people at several institutions. Eventually, grants are made to the individuals involved, usually for longer periods of time than under the general grant-making programme.

Last year, the foundation launched seven new programmes, committing SFr 12 million each to "artificial intelligence" and the "chemistry and physics of surfaces". A programme on "human health and the environment" is supported with SFr 14 million and a study of the future of cities and their transport systems with SFr 12 million. The remaining projects deal with socio-political questions.

There seems to be general agreement that these programmes would be more effective if they could be more quickly formulated and financed. The decision-making process, entailing as it does iterative approval by the subcommittee responsible, by the council of SNSF and by external bodies, is generally agreed to be slow. Some of the projects eventually approved are also somewhat amorphous in character. The SFr 16 million programme on "materials for tomorrow's

needs", for example, appears to be a collection of projects in materials science that may each be estimable in its own right but which, collectively, appear largely to suggest Swiss interest in the encouragement of an important field of research, and in keeping in touch with research elsewhere. Both these criticisms of the National Research Programmes were raised in 1985 by an independent assessment of the schemes by D. Freiburghaus and W. Zimmerman (Paul Haupt, Berne; 1985). But it is unlikely that SNSF will be able, simply by taking thought internally, to change the pattern of this part of its work. Foundations, after all, can do no better than the applications they receive from their constituents, in this case the company of Swiss researchers. In the last resort, researchers themselves may have to challenge SNSF with projects that break new ground.

J.M.

Office of Education and Research

Pushing at an open door

"WHEN I ask the Parliament for money for a new project, they usually say, 'Are you sure that's enough?'" Professor URS Hochstrasser, director of the Office for Education and Research for the past 22 years, does not by that imply that the Parliament does not know what to do with the cantons' money, but that he is too skilled at his job to make immodest demands on public funds.

The result has been a steady stream of new projects produced by midwife Hochstrasser. His office, for example, found the funds for the disproportionately large Swiss contribution to the L3 detector (see page 328) now being installed at CERN's new electron accelerator, LEP, at Geneva. (Switzerland will pay a sixth of the total cost of SFr180 million to accommodate the enthusiasm of a Swiss group led from ETH, Zurich.)

Similarly, Hochstrasser's office has pushed hard so that Swiss universities can have access to modern computing machinery. He reckons that he has found SFr207 million for hardware in the past few years, of which SFr40 million has been spent on supercomputers (at Lausanne) and on giving academics access to them. The scheme for moving the plasma physics group to Lausanne similarly began in the same office.

In the Swiss fashion, and despite the formal hierarchy of the advisory committees, it seems that anybody can put forward a proposal for big spending. But prudent entrepreneurs take care that the schemes they advocate carry the approval of their colleagues, who may be members of other committees, in the line of consultation, with the result that even the most ambitious schemes are whittled down to manageable size.

The plan to fuse together the Nuclear Reactor and the Nuclear Physics institutes into the Paul Scherrer Institute was, says Hochstrasser, an initiative of the Science Council (which advises the Federal Council) which he nevertheless believes to have been well-justified; at a time when the role of each institute must change, why not also honour Paul Scherrer not only for his X-ray work but also for having constructed a cyclotron at ETH Zurich as early as 1939.

But why does Switzerland have to be engaged in nuclear physics and reactor technology as well as in CERN? Hochstrasser says: "Well, the Dutch have astronomy, do they not?" His point is partly that small countries must somehow find ways of making a mark internationally, partly that doing just that powerfully encourages other research groups. To do that, it is necessary to back outstanding people with sizeable sums of money even if their projects appear to have no immediate usefulness.

The other side of this coin is the Swiss interest in overseas collaboration with organizations as different as the European Space Agency and the EUREKA programme. One obvious objective is to keep in touch with emerging technology, but at least as important a consideration is that Switzerland should earn a high reputation in these collaborations both as a mark of its genuinely international interest (best indicated by the large proportions of foreign students at all Swiss universities) and in case there should be a future need of friends; even in the middle of Europe, a small country can accumulate a sense of isolation.

There are more immediate problems in the front of Hochstrasser's mind, the chief

of which seems to be that of the acute shortage of technical people experienced by Swiss industry. Both in the universities and polytechnics, there are too few students following courses in chemistry and physics, engineering and computer science, he says. The confederation's only weapon is "counselling", another word for persuasion. One guesses that the first Swiss to find a plausible way of doing that effectively will have Hochstrasser jumping at any offer of extra funds the members of the Parliament may make.

The other academic problem in the front of his mind, echoed in the universities, is that the professoriat has become static. Hochstrasser says that a quarter of university professors have been in their posts for twenty years or more — a problem familiar elsewhere in Europe. One

difficulty, in Switzerland, is that buying flexibility with schemes for early retirement would be fearfully expensive, given that a professor's basic salary is SFr150,000 a year, the equivalent of £57,000 or of \$103,500 a year.

There is a smaller battle yet to be won. Perhaps the outstanding anomaly of the Swiss universities is the circumstance that the academic years of the German- and French-speaking universities do not coincide on the annual calendar. The University year begins in April at the University of Zurich, for example. Hochstrasser believes he may have won a battle for rationalization of this pattern, but his countrymen will suspend belief until they are sure that nobody has raised the 30,000 signatures required to call a referendum on a proposal for change. **J.M.**

Geneva University

Still a struggle to keep ahead

NEXT to Basle, the University of Geneva is the oldest of the Swiss universities. (Founded at the end of the sixteenth century, it was a Lutheran stronghold from the outset, and still has a Department of Protestant Theology.) But the dean of science, mathematician J.-P. Imhoff, is to be found on the fifteenth floor of Geneva's television tower when not practising as a mathematician.

Inevitably, jammed against the border with France, Geneva is more cosmopolitan even than other Swiss universities. Imhoff reckons that a third of the present faculty is foreign, but says there is now pressure on the university to appoint more Swiss nationals.

There has also always been a cosmopolitan tradition among the student body — Russian refugees at the turn of the century, for example — and some 30 per cent of the students at the university are now non-Swiss, mostly French. Roughly 40 per cent of science students are from outside Switzerland. Geneva is also distinguished from most other Swiss universities in that over half of its students are women.

But the science faculty (with 1,800 students, 450 of them graduate students of various kinds) is smaller than the faculties of letters and, in particular, the faculty of economic and social sciences, which has more than doubled in size in the past decade to a total of nearly 2,500 students.

Among the science departments there are some, especially in biology and chemistry, which are short of students. The consequence is that academics enjoy luxurious teaching loads and have often ample time for research. The budget of the university from the canton of Geneva, over SFr 250 million a year, is sufficient for there to have been an improvement in the quality and condition of research equip-

ment during the past four years.

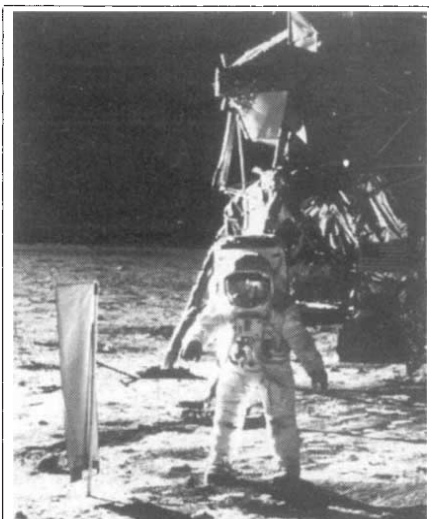
Among other things, Geneva's research reputation in the past decade or so turns partly on the success of its molecular biology, largely because of the department of molecular biology founded by A. Tissières, who retired at the end of the last academic year (and who has been succeeded by Professor U. K. Laemmli).

But in a closely related field, the university has also been hugely embarrassed by the accusations of scientific fraud levelled in 1983 at Professor Karl Illmensee, a member of the embryology department at the medical school.

Imhoff says of the affair, which culminated in a critical report by an independent commission appointed by the university and in Illmensee's resignation in July, 1985, that the "dust has now settled", but he also acknowledges that the incident "has left marks". The report of the committee of inquiry was critical of the university for indecisiveness, and for its failure properly to confront Illmensee with the charges against him.

One of the more obvious lasting marks may be the reorganization of biology teaching now brought about by the creation of a new department of biology, the effect of which seems to have been to rob the molecular biology department of two and a half senior posts. There are some who suspect that cutting down molecular biology in this way is delayed retribution for the department's support of Illmensee's original appointment.

Another mark of the affair is the damage done to Illmensee himself. Laemmli sticks by the view that he is a man of great talent who has now been excluded from professional life. He says of those found guilty of misdemeanours in research that we "expect them to pay a very heavy price". **J.M.**



JOHANNES Geiss, from the Physics Institute at the University of Berne, has built up a considerable space research facility over the past two decades with substantial support from SNSF, but is in many ways most proud that Neil Armstrong and Edwin ("Buzz") Aldrin took one of his early experiments with them on their first journey to the Moon.

The intention was to collect particles from the solar wind on a platinum foil screen supported on the surface of the moon as if it were a flag mounted on a stick stuck into the ground. The screen was returned to Berne, but the support remains where it was placed. Similar equipment was carried on five succeeding Apollo flights.

Geiss has used the measurements to advance an interesting hypothesis — that the proportion of deuterium in the nebula from which the Sun condensed must have been less than the proportion now found on the surface of the Earth. The argument is that the ³He measured in the solar wind should consist of the proportion with which the Sun was originally endowed together with that created in the history of the Sun by the conversion of deuterium. The notion, which emerges, that there must have been fractionation processes in the early Solar Nebula which have increased the proportion of deuterium on the surface of the Earth above the primordial value.

Meanwhile, Geiss's enterprise has grown from being a means of mounting modest experiments (the Apollo foils, sticks included, weighed merely 250 grams) to a near-industrial enterprise with a payroll of 120, the large technical staff included.

Its objective is to work up experiments that can be carried by other people's satellites and rockets, usually in collaboration with other people. At the time of the Halley encounter in 1986, for example, he and his colleagues helped design and analyse the data from an ion spectrometer carried by the Giotto spacecraft.

But now, Geiss presents a problem for his sponsors by his impending statutory retirement, which will require some part of the Swiss government to decide whether the future of his enterprise should rest with the University of Berne and grants from SNSF, or become an independent institute, with support from some other part of the government. **J.M.**

Centre for plasma physics research

Progress towards fusion

How does Switzerland, jealous of its national independence, manage nevertheless to behave as a member of Euratom, the European Community's agency for the promotion of nuclear energy? That is how things are with the Swiss programme of research in thermonuclear fusion, embodied in a neatly engineered tokamak machine at the Centre for Plasma Physics Research (Centre de Recherches en Physique des Plasmas), on the campus of the federal polytechnic at Lausanne.

The Swiss logic is undeniable: a country ever-conscious that unexploited sources of hydroelectric power are physically limited and certain to be more expensive than those now in use cannot stand aside from the development of potentially novel sources. So, like Sweden but later, the Swiss confederation started making suppliant noises in Brussels in the late 1970s and became a formal member of the Euratom thermonuclear project ten years ago. Swiss now participate in work at the Joint European Torus at Culham in Oxfordshire, and the confederation sustains an impressive domestic programme.

The tokamak has no pretensions of being a machine in which fusion will actually occur; it is, rather, principally a device for generating plasma to be used for developing diagnostic and heating tech-

avoiding contact with the containing walls — a notorious thief of temperature and source of impurity.

The visual proof of this intention is a wooden mock-up of the intended cross-section through the torus and its surrounding electrical coils — the first next step will replace the present circular tube of 18-cm radius with an elliptical cross-section taller (at 150 cm) than it is wide (36 cm), but this is also why the machine is known not only as TCA (where 'C' stands for '*chauffage*' and 'A' for 'Alfvén'), but as TCV (where 'V' means '*Variable*').

Like other distinctive features of the Swiss programme, the design of the tokamak has been coordinated within the Euratom programme. These arrangements, while infringements of people's freedom of action, bring benefits as well. Euratom normally pays for 25 per cent of the cost of participating thermonuclear

programmes, but 45 per cent of those judged to be distinctive contributions towards the common goal.

The Lausanne centre boasts others, of which the most intriguing is the laboratory's exploration of the use of Alfvén waves (after the Scandinavian physicist Hannes Alfvén) as a means of heating plasmas. The basic functioning of now-familiar tokamaks depends on the principle of a transformer. By driving an electrical current around circuits enveloping magnetic material and a torus containing ionized gas, the conducting medium inside the tube functions as if it were the secondary winding of a transformer. So it will carry an axial electrical current so long as the current in the external coils is increasing, and will be compressed or 'pinched' as its own current increases.

The snag is that even when the electrical engineers have done everything they can to maximize the energy loaded into the plasma by this means, spontaneous thermonuclear fusion will not persist for long enough to be economic (and, in any

Swiss making waves at CERN

THE L3 detector at the LEP electron-positron collider ring at CERN near Geneva is the biggest international collaboration in the history of high-energy physics — a partnership among 400 physicists and 400 engineers and technicians from 34 universities in 13 countries, including China, the Soviet Union and the United States.

Building the detector will cost SFr200 million; the machine should be counting its first particles soon after the JET ring, 27 km in circumference, has been commissioned in the middle of next year. The detector is being installed at a point on the circumference at which the counter-propagating electron and positron beams interact, and has been designed to record the passage of as many as possible of the particles produced in the electron-positron collisions which are the objective of the machine.

Hans Hofer, a Swiss physicist from ETH Zurich, who has helped oversee the project, says that Swiss support has been invaluable. The Swiss government, he says, "likes international collaborations" and is contributing an unprecedented SFr31 million to L3. Hofer explains that L3, one of four experiments at LEP, will measure just a few elementary particles at very high precisions.

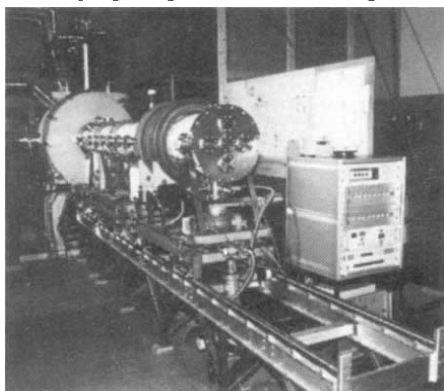
The particles produced in the electron collisions will include both electron-like leptons and hadrons, or strongly interacting particles. Hofer says that the aim in the design of L3 has been to measure the characteristics of the electrons and photons with a precision of one per cent, ten times better than achieved with existing

detectors. By logging all the particles emitted from an electron-positron collision and fitting streaming chambers near the axis of the beam to record any novel particles produced, L3 should make it possible to reconstruct the complete roster of particles formed in electron-positron collisions outside the range of energy so far explored.

Hofer attributes the success of L3 in raising funds to project leader (and Nobel laureate) S. C. C. Ting's connections in the governments of participating countries, from the People's Republic of China to Bulgaria to the Soviet Union. L3, says Hofer, is "much like the United Nations, only it works".

One direct benefit to Switzerland from L3 will be the experience of building a machine on this scale. Another is the training of Swiss physicists. Dozens of Swiss students are among the one hundred doctoral candidates at L3. Although their training will have been specialized, which could be a drawback in trying to find a university position, Hofer says that they will have "ready-made contacts" at 34 universities around the world.

The indirect benefits are more difficult to assess, but may be even more substantial. The effect of the many international research organizations based in Switzerland on the temper of Swiss science cannot easily be gauged, but is probably much less than might be expected. Part of the trouble is that researchers at institutions such as CERN are often transient. It will be interesting to see whether the unusually intense collaboration offered by L3 will have more lasting effects. S.D.



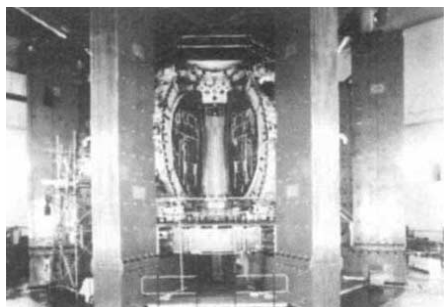
Reason to boast — the gyrotron at Lausanne.

niques, but which has been designed deliberately so as to be useful in testing other than circular cross-sections of the toroidal tube in which the plasma is contained and heated. (The trick is to build the torus and its surrounding coils so that the top half can be lifted off, allowing the replacement of a different torus.)

This part of the project, officially part of the Euratom programme since 1985, is reckoned to be potentially important because asymmetrical plasmas may turn out to be more efficient means of generating the combinations of high temperature and density at which fusion will be spontaneous, or they may be better suited than circular cross-sections to keeping plasmas stable or, more practically, better at

case, will be made more economic if the persistence can be increased). As things are, the addition of energy by means of beams of neutral hydrogen particles is favoured as the means of adding extra energy, but Alfvén waves are another way.

Appropriately, this mode of electromagnetic wave transmission was first recognized in connection with astrophysical plasmas, which is why Alfvén waves are most commonly heard of in connection with transmission through the solar wind. The phenomenon is a kind of generalization of that in which the plane of polarization of a light wave will rotate during passage through an ionized medium; in a tokamak filled with plasma with uniform charge density, it should be possible to set up along the axis helically rotating standing waves, but when (as in real life) the charge density is not uniform, the frequency will be a function of the radius and usually greatest at the outside (introducing a progressive shear of the plasma). Or,



Collaboration — ABB produced the toroidal and poloidal field magnets for the tokamak reactor at JET, UK.

turning the argument around, an ionized plasma should absorb high-frequency electromagnetic energy from an Alfvén preferentially at some fixed radius, whence the hope that this might prove a way of prolonging the high-temperature regime of a hot plasma.

What the Swiss at Lausanne most boast about, however, is their way of generating energetic pulses of electromagnetic energy in the frequency range above 10 GHz. Their device is called a gyrotron — a way of coupling through an appropriate magnetic field the energy of an electron beam with an electromagnetic wave in a resonating cavity. The Swiss device differs from others in that the resonating cavity is perpendicular to the electron beam. So far, the Lausanne centre has built a tube producing 500 kW of 8 GHz radiation in 10-second pulses, but there are hopes of reaching frequencies of 150 GHz. Asea Brown Boveri (see page 338) is closely involved, as might be expected.

As plasma physics laboratories go, that at Lausanne differs from those elsewhere in the world mostly in the respect in which Switzerland itself differs from the other places: the laboratory and its machines are smaller, but their quality is as good, perhaps better. Naturally, nobody at Lausanne

will at this stage say when, or whether, fusion will be a technical reality which is also economically practicable, but the cautious raising of excitement apparent elsewhere is shared in Switzerland.

But will it not be an especially serious disappointment for a country in which everything is conditioned by its smallness if fusion should, in the end, come to nothing? The general opinion is "No". For one thing, there are tangible spin-offs, such as the gyrotrons that will be made for other purposes, even if high-frequency heating does not push thermonuclear fusion over the edge of economic feasibility. There is also the inherent interest of the difficult problems with which plasma physicists

grapple all the time, and whose solution cannot but help people somewhere.

But the way in which the Swiss fusion project has won entry to an international club — not just Euratom's own club, but the international pool in which that institution swims — may in itself be worth all that effort and even the money that sustains it. For all its technical excellence, the Swiss fusion project typifies Swiss views of the big wide world away from the mountains. The cantons may be wrapped up in their own domestic and even smaller problems, but the confederation to which they have delegated such limited powers is, as always, alive to the need that some Swiss should play on a bigger stage. J.M.

Nuclear energy

Power to the reluctant people

SWITZERLAND, which was toying with the notion of making nuclear weapons to defend its neutrality, may now be on the point of abolishing nuclear power stations.

A de facto moratorium on the construction of new plants has left the Swiss government wondering how it might replace the more than one-third of Swiss electricity now provided by nuclear power. Though a referendum was not the catalyst for the decision, the unique Swiss system of direct democracy ensures that the government cannot force an expansion of the nuclear programme on an unwilling population.

Before the 1986 Chernobyl accident, Switzerland had approved sites for three new nuclear power plants and was considering building up to three more by 2025. The anti-nuclear movement had failed in a 1984 referendum to halt all new facilities. The referendum lost by 55 to 45 per cent. But Chernobyl led to a reassessment. In a shocking change of heart, the right wing of the Swiss parliament withdrew its support from the most advanced of the new reactors, a 1,000-MW plant at Kaiseraugst near Basel. The government will have to forfeit about one-third of the SFr1,200 million already invested. Politicians from Berne and Geneva immediately followed with requests to abandon plants near those cities. Graben, near Berne, was to provide 1,000 MW and Volbois near Geneva, 330 MW. Although the government has the option to resume work on these reactors, observers say they will never be built.

About 37 per cent of Switzerland's electricity comes from the five operating nuclear plants. Most of the rest is generated by water in the Alps. This hydroelectric power is also sold at a profit to other countries during peak periods or exchanged for base-load electricity supply. But the amount of additional hydropower that can be obtained is limited to 15 per cent by environmental considerations.

An official in the Swiss Federal Energy Office, Martin Renggli, outlined the "political boundary conditions" restricting government action. The government rejects relying on electricity imports from France (which generates its power primarily through nuclear plants). It does not wish to increase dependence on imported fossil fuels, and Switzerland has no significant oil or coal reserves.

Without an increase in nuclear generating capacity (which admittedly brings environmental concerns of its own), the only viable long-term solution is energy conservation. But the government's power is limited by its lack of authority under Switzerland's federal system. It will attempt to bring about a constitutional amendment to allow it to mandate conservation measures in the cantons.

The short-term solution has been to sign an agreement with neighbouring France to buy 1,550 MW of power between 1989 and 2015. But the Swiss government is determined to keep the nuclear option open for the future by supporting nuclear research (with up to half of the federal energy research budget) and by maintaining existing plants for as long as possible. Eduard Kiener, director of the Federal Energy Office, calls it "politically dishonest" to shut Swiss nuclear plants and then buy the energy of one-and-a-half nuclear plants from France. Furthermore, the supply is not guaranteed against strikes or political shifts. Meanwhile, opponents of nuclear power are pressing ahead, encouraged by the uncertainty and division shown by proponents.

Two referenda are coming up in 1991 at the latest. If passed, each referendum would become part of the constitution, and parliament would be required to implement it. The first proposes a moratorium on the licensing of any new plants for ten years; the second calls for the phasing out of nuclear plants "as soon as possible". S.D.

Snow and avalanche research

All you ever wanted to know about snow . . .

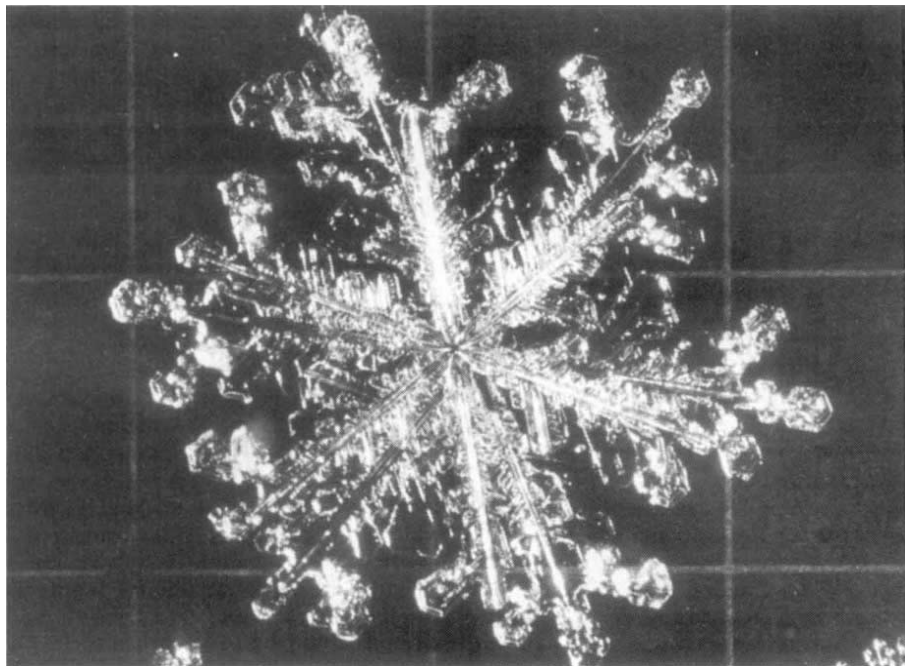
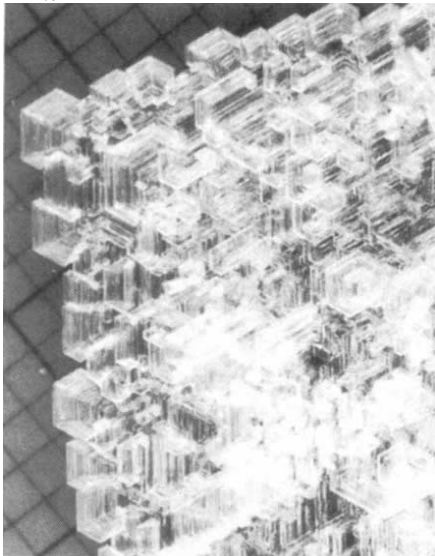
"THERE are as many types of snow as there are of wine", says Claude Jaccard. He ought to know — Jaccard has spent most of his professional life studying snow and ice in the Swiss Alps, and he seems to have found the perfect situation as director of the Federal Institute for Snow and Avalanche Research (SLF is the German acronym for Schnee-und Lawinenforschung).

Perched near the top of an alp called Weissfluhjoch near the resort of Davos in eastern Switzerland, the institute overlooks a narrow, V-shaped valley. At 2,760 m, there is snow six months a year. A staff of 44, including 6 physicists and 11 engineers, commutes to work daily by the Parsennbahn cog railway.

Research at the institute has come a long way since 1936, when it was established in a small building near its present site, a warm, well-equipped stone edifice at the upper terminus of the Parsennbahn. The number of people killed in their homes by avalanches has decreased dramatically in that time, in part because of a sophisticated warning system administered by SLF. The number of skiers and other tourists killed by avalanches has increased threefold over 48 years, but the overall number of tourists has increased "50- or 100-fold", says Jaccard.

SLF is the oldest institution of its kind in the world. Although most of the other alpine countries as well as the United States and Canada have similar institutes, SLF is unique in the mixture of research and consulting undertaken. From next January, it will be affiliated with ETH in Zurich and no longer attached to the national forest service.

Avalanche prediction and prevention is large fat crystals have metamorphosed into 'depth hoar', a layer of snow disposed to slide. The grid is 2 mm.



A 'double crystal' of freshly fallen snow. This photo and the two below from SLF.

a mixture of modern and traditional technologies. Tracked vehicles carry radars across the snow-pack to determine the structure of subsurface snow.

SLF researchers even plan experiments with a form of sonar, which involve dragging loudspeakers and microphones out onto the snow, as a way of avoiding electrical interference that confounds the radar equipment.

But volunteer observers use such commonplace items as a metal pole, a scale and a jeweller's loop for their daily measurements of the height and density of the snow and its crystal form and size. The volunteers submit reports twice monthly on the snow pack stratification and its resistance measured with a Rammsonde or 'ramprobe'. The data can reveal clues about impending avalanches.

Although all the research at SLF is done with applications in mind, the staff probes the physical world with as much curiosity as any university research group. Samples from the extensive photographic archive as well as charts and graphs adorn the hallways, demonstrating the results of experiments or the catastrophic outcome of avalanches. The most beautiful of the photographs are those of snowflakes in unimaginable variation.

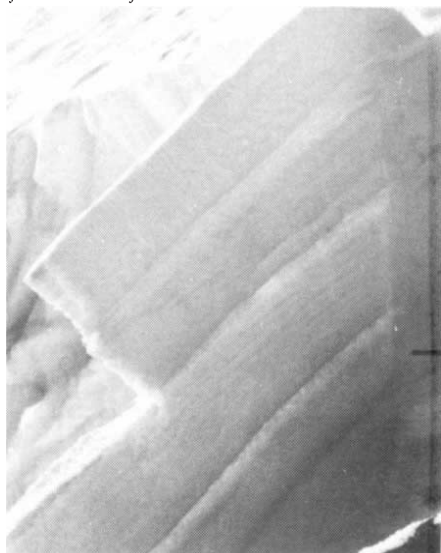
The individual snowflake — or 'grain' — is where an avalanche begins. "What the atom is to matter, the grain is to snow", says Jaccard. Snow on the ground undergoes a metamorphosis during the course of the winter as the grains establish new forms and relationships. At first, there are the familiar star-shaped crystals exhibiting fine detail (the photograph

shows a double-crystal flake).

Then the details disappear, leaving small, rounded crystals. The crystals may grow larger with time, developing well-defined faces and edges. When large sheets of coherent, flat crystals form, it means *danger*. This is 'depth hoar', the immediate possible precursor of avalanche conditions. One thin layer of depth hoar is often the weak link (see photograph) that can send thousands of tons of snow crashing down the mountain.

Jaccard, a native of the mountain town of St Croix in western (French-speaking) Switzerland, is an old hand at snow research, having done a physics PhD at ▶

This is where an avalanche starts, the thin layers of 'depth hoar' providing the weak point when further snow falls.





SLF director Claude Jaccard: a three-fold increase in deaths due to avalanches has to be seen in the context of a "fifty-to-hundred-fold" increase in tourism. On the right, avalanche barriers on the lower slopes of Weissfluhjoch, Davos.

ETH on the dielectric constant of ice. He returned to Switzerland after two years at Argonne National Laboratories in the United States and went directly to SLF for a six-year stint. He returned as director in 1980 following 13 years as professor of solid-state physics at the University of Neuchâtel.

Understanding the mechanics of snow is difficult, says Jaccard, because its structure is disordered. Whereas describing the mechanical properties of a newtonian liquid requires just a single parameter, viscosity, describing the mechanics of snow, requires ten or fifteen. In one continuing project at EISLF, researchers take snow samples, make thin sections of them,

fix and photograph them against a dark background and analyse the photos by computer. They are now experimenting with three-dimensional computer models to improve accuracy.

SLF has already contributed to new zoning laws and architectural norms. Communities heed its warnings and evacuate or alert their residents depending on the danger. Long-term efforts, for example by planting trees or erecting preventive structures (shown above), continue to reduce the danger of avalanches.

Such measures have their own ecological consequences. Trees must be planted all the way to the top of a ridge and must be planted densely enough to intercept

snowfall efficiently.

The trees themselves are often the first victims of avalanches when this plan is not followed. Small trees may bend under the force of the snow but larger trees are mostly destroyed.

Beneath the man-made structures, avalanches occur less frequently, but the snow up above melts more slowly, shortening the vegetation period. Some plants below the barriers get a kind of frostbite when not covered with a blanket of snow.

Facilities at the institute include room-sized refrigerators used to study snow. Asked if the researchers save snowballs for use in the summer, Jaccard says, "No, we see enough snow in winter". S.D.

Science faces a struggle for popularity after Schweizerhalle

CONCERN about the side-effects of science is at an all-time high in Basle in the aftermath of the 1986 toxic chemical spill at the Sandoz facility at Schweizerhalle. Coming after the dioxin spill at Seveso, Italy (for which Hoffmann-La Roche was responsible) and Chernobyl, Schweizerhalle was "the last straw", says one researcher. For the first time, residents were confronted directly with an environmental catastrophe, in which most of the fish in the Rhine died.

In the most ominous development, windows were broken at the Biozentrum of the University of Basle, as if basic researchers were to blame for the spill. Opponents of animal research and genetic engineering have taken advantage of the ferment to push restrictions in those areas. Although no laws have been passed, some researchers fear that the next proposals will be "more reasonable and therefore more dangerous". And concern is growing about anti-science sentiment elsewhere in Switzerland as well.

Opinions on the matter vary. Peter Fricker, general secretary of the Swiss National Science Foundation (SNSF), says that most Swiss have a tolerant, even a "benevolent", attitude towards science, if they care about it at all. The National Research Programmes of SNSF serve as a "locomotive" to help create a positive public image of science.

Alfred Pletscher, president of the Swiss Academy of Medical Sciences and long-time science administrator, says that the worst of the backlash from Schweizerhalle has passed. Parliament has been supportive toward science for much of this decade, because "they realize that research is necessary to keep a country with few natural resources strong". Fricker sees the European Community "fusion" of 1992 as the kind of threat that could unify the population behind science.

There has been an effort, says Fricker, to publish more about science in daily newspapers to "make people see the good side of

science". Eckart Gwinner of Hoffmann-La Roche sees a need for more and better training of science journalists.

Basle biologist Eduard Kellenberger takes a more pessimistic view. He predicts that the debate over the merits of science will become "emotionalized and polarized" in Basle much as in West Germany. Opponents of genetic engineering, for instance, refuse even to meet university or corporate researchers. A parliamentary proposal to outlaw genetic engineering remains shrouded in secrecy.

Hoffmann-La Roche and Ciba-Geigy have increased their public relations efforts in the form of public meetings and open houses at factories and laboratories to try to keep open the lines of communication. Forums have been well-attended but the battle for hearts and minds has only just begun. Kellenberger, for one, thinks that it is already too late. "No matter what the companies do, the population doesn't believe a word." S.D.

Federal Institute of Technology

Cosmopolitan route to excellence

THE first professors at Switzerland's high-powered Federal Institute of Technology (Eidgenössische Technische Hochschule or ETH) were refugees who fled prison sentences or political repression in Germany after the failed 1848 rebellion. They found a haven in Switzerland, where the government was building a polytechnic-cum-university to train engineers, architects, science teachers and other professionals.

Many current ETH professors are refugees of a different sort. Lured away from top institutions in West Germany, the United States and Britain, they seek shelter from the struggle for grant money that prevails elsewhere. The teaching load is light, the students generally well-behaved, and, as one US researcher put it incredulously, "every chair is endowed".

ETH aspires to being "one of the twenty best universities in the world" in teaching, research and international reputation, in the words of its President Hans Bülmann. Nearly half its faculty are foreigners. There have been three Nobel prizes granted to ETH professors and ten more Nobels for researchers who had studied or worked at ETH but did their prize-winning work elsewhere.

The willingness to hire on merit rather than nationality has helped ETH succeed at its original mission. The accent on

excellence has arguably helped small, resource-poor Switzerland to become one of the wealthiest countries in the world in terms of per capita income.

One meets foreigners in every stage of their careers at ETH, from graduate students to professors emeritus, by which time many have become naturalized Swiss. Biophysicist Tim Richmond, a US citizen, exemplifies the type of researcher ETH seeks. ETH lured him away from the MRC Laboratory of Molecular Biology at Cambridge (UK) primarily because it could offer him the chance to expand his group.

Richmond and his colleagues had determined the crystalline structure of the nucleosome core particle at the MRC (T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes and A. Klug *Nature* **311**, 532–537; 1984), a substantial advance. He wanted to try his hand at several new projects, some of which (like attempting to crystallize the DNA-binding region of glucocorticoid receptor bound to the nucleosome) required long-term support. ETH promised him the resources he needed with "the fewest administrative problems of any university I looked at", the others being top universities in the United States.

According to organic chemistry professor Jack Dunitz, a Scot, the average grant proposal in Switzerland is four pages

long, and the average chemistry professor spends five or six days a year writing grant applications. At the US National Institutes of Health, by contrast, people spend "50 or 60 days" on the task. The chairs are endowed in the sense that two-thirds of a professor's operating budget is provided by the university. The rest must come from outside sources, in most cases the Swiss National Science Foundation.

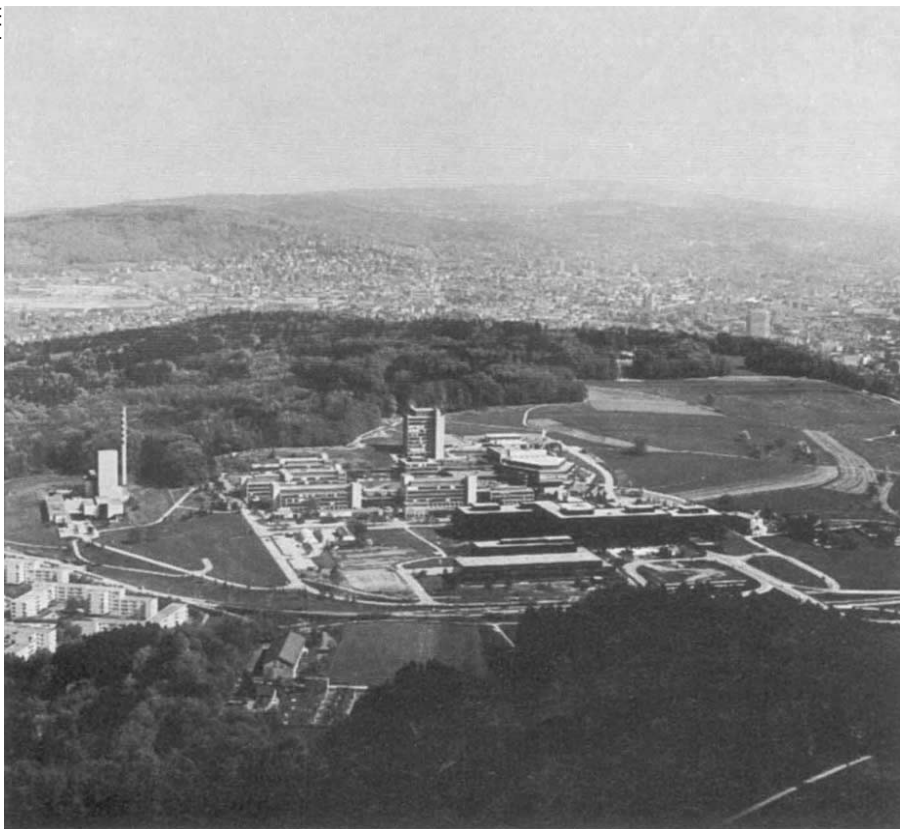
The main difficulty in moving to ETH? The language. High German is required, but "nobody speaks it — they all speak Swiss-German", says Richmond. The two are almost as different as, say, English and French, as anyone who speaks fine high German can attest after a (frustrating) trip to Switzerland. Holding lectures in English for a year or two is all right, but eventually German is a must.

Persistent probing reveals that there are some real difficulties at ETH. Some problems are directly related to success — shortage of space, for instance. ETH's central Zurich campus is almost completely saturated, and expanding into the ever-green forest just uphill from the outlying Hönggerberg campus is not as easy as it looks.

The abundant funds at ETH's disposal sometimes lead to money being invested without enough forethought; the super-computer installed this year is one example. Professor Niklaus Wirth, the inventor of the Pascal programming language, fears that ETH administrators may be disappointed in the results of a five-year investment programme for new equipment. Wirth also resents the image of information sciences as some sort of service department, expected to churn out lots of graduates to fill a gap in the economy without contributing significantly to research. But he has never been prevented from doing good research in Zurich, and on sabbaticals.

Another, more deeply rooted problem is complacency. Though most professors denied that this is a big problem, President Bülmann echoed John Kennedy in addressing students and faculty members last year: "Ask not what your future at ETH will give to you; ask what you can do for the future of ETH".

Ralf Hütter is the first 'vice-president for research', a position created last month. His main task, he says, will be "to create a supportive environment for research". Having someone responsible just for research will give professors greater access to the administration and, he hopes, will help them to identify more with the institution. "At some schools, you are a Harvard professor", says Hütter. "Here, we have traditionally been professors [who happen to be at ETH]." The largest task for ETH in the coming years will be to turn out more engineers. There is already a severe shortage in Switzerland which shows no signs of abating. S.D.



The ETH campus at Hönggerberg — seen from the air in 1980.

Academic living

THE most striking feature of Swiss academic life is that its practitioners are well paid. Salaries of professors begin at about SFr150,000 a year, comparable with those paid at all but the most well-heeled of universities elsewhere in the world.

The Swiss say that there is no alternative: the universities have to be able to recruit people from where they choose. What this implies is that the cantonal universities must be in a position to recruit people from universities elsewhere, while the two polytechnics must be able to attract successful engineers (more often Swiss) from successful industry.

More than that, appointment as a professor at a Swiss university carries with it not merely prestige and an endowment in equipment but ample opportunity for research. And at least some cantons are prepared, on behalf of senior academics, to relax the usual rules preventing incomers from buying houses and apartments without having lived in Switzerland for five years, or their spouses from taking paid employment.

At least in money terms, much the same applies to more junior academic scientists. At the University of Geneva, for example, a teaching assistant (appointed for a maximum of five years) could expect to earn SFr4,000 a month, say SFr50,000 a year, while preparing a doctoral thesis. This is in accord with what seems to be a general principle of Swiss life, that the disparity between the salaries of senior and junior people is less apparent than elsewhere.

Can there be, in such halcyon circumstances, discontents? Sadly, but inevitably, there are. The most common is that the many people occupying university posts below the rank of professor or those on postdoctoral fellowships foresee only a bleak future. On one estimate, there may be 10 able scientists in the system for each chair likely to become vacant in the next 15 or 20 years. But the prospects of promotion are made worse by the inclination of Swiss universities to recruit academics from elsewhere, a potentially bruising issue which remains unresolved.

There are also worries about the administration of the universities. No canton has more than one university (and most have none), with the result that the institution tends to be regarded by the canton government as its department of higher education. The result of that is a degree of closeness in the relationship between universities and cantons that requires the university administration to present a bureaucratic face to the outside and thus, internally, to be aloof — sometimes high-handed in its internal dealings, sometimes over-cautious. It is perhaps as well that "freedom of research" is guaranteed by the law. J.M.

University of Zurich

University education for all?

RESIDENTS of Zurich take higher education seriously. In 1970, they voted to spend SFr600 million to build a new science campus on farmland just outside the city centre. Until recently, they spent a remarkable 10 per cent of the canton's budget on the University of Zurich. The figure has dropped to 8 per cent, but even University Rector Hans Schmid, a theologian, admits "there have to be some limits". Zurich is the largest of Swiss universities, with almost 20,000 students. Between the University of Zurich and ETH, a quarter of all those enrolled in Swiss universities study in Zurich.

But, although ETH has a higher profile in the world at large, there is no enmity between the two institutions: they share facilities in a number of disciplines and make common arrangements for dormitory space and the like. "We look at them like a younger sister", says university secretary Maximilian Jaeger. (The university was founded in 1833, 22 years before ETH.) The science campus at Irchel is nearly complete. The third and last section should be ready for occupation by molecular biology, physics, mathematics and other faculties in March 1989. One molecular biologist called it "astounding" that the university could put together enough land so close to the city centre. Once the move is complete, the city centre campus, adjacent to ETH, will house only humanities and social sciences.

The university has been subject to tremendous growth — the number of students has increased fivefold in 25 years — in part due to a quirk of the Swiss university system. Unlike in West Germany and other countries, there are no quotas to limit the numbers of students in particular fields. Even in medicine, anyone who wants to study can do so for a nominal fee: the number of students has surged to 2,600 in medicine and dentistry. But by and

large a good education can be had by all — there is a mechanism to divert excess medical students to other universities.

Natural scientists have remained mostly unaffected by the student boom. Noted molecular biologist Charles Weissmann, a Swiss native who returned from New York University in the early 1970s, spoke of "generous" support and good facilities. Not so in the humanities and social sciences, where Schmid says "education has suffered" because of the overload.



Main entrance to Zurich's city centre campus.

Student/faculty ratios are up to 120 to 1 or even 150 to 1 in economics and psychology. The law and German departments are also overburdened. Research has necessarily been pushed aside to make room for the extraordinary teaching load.

The university administration hopes the numbers will stabilize in the coming years. Jaeger says "we hope that the big problems are behind us". This is not inevitable despite the passing of the "baby boom" generation. The student body is changing demographically, with more women studying and older people returning for second degrees. And some cantons are still striving to increase their educated population — an interesting phenomenon in a country that appears uniformly wealthy and successful to a casual visitor. S.D.

Basle University

A victim of local tensions

BASLE has a small university with a very long history. Founded in 1460 by a papal bull from Pope Pius II, Basle predates Lausanne by more than 70 years as the oldest Swiss university. As Enea Silvio Piccolomini, Pope Pius II had been secretary to the Basle Council from 1431 to 1449. Residents clamoured for the honour of a university (which in those days had to be created by the pope or an emperor) and the pope granted their wish. Even then, there was dissent among Basle's citizens over the advisability of such a large and expensive undertaking.

Five centuries later, support for the

university remains a politically hot issue in the divided canton of Basle. Some historical background is in order. As in other Swiss cities in the nineteenth century, war broke out in 1833 between the city of Basle (Basle-Stadt) and its hinterlands (Basle-Landschaft) that led to a partition. A division of the university's worldly goods similar to a modern-day divorce proceeding was avoided only through the intervention of the Swiss Confederation and the payment of a ransom by Basle-Stadt to Basle-Landschaft.

Basle has lagged behind other cities in resolving the country-city differences.



Some institutes of the university are in buildings of the fifteenth and sixteenth centuries.

This did not present a threat to the university until the late 1960s, when the fat years suddenly grew lean. The budget began to shrink and Basle-Stadt sought support elsewhere.

Basle-Landschaft, which provides one-third of the students at the university, agreed to pay about one-fifth of the costs. It might have agreed to pay more if it had been given more say in university affairs. But there has never been much trust between the two half-cantons. As one professor puts it, "each side tries to blackmail the other". In a referendum, Basle-Landschaft overwhelmingly rejected a proposal to reunite the split canton a few years ago, fearing domination by the inhabitants of Basle-Stadt.

Today, attracting more support from Basle-Landschaft, the Swiss federal government (which already makes a sizeable contribution) or some other canton is the university's main concern. The 200,000 inhabitants of Basle-Stadt do not provide a large enough tax base (even with help from the three pharmaceutical giants Ciba-Geigy, Hoffmann-La Roche and Sandoz) to keep the university, especially its medical clinics, going. (The clinics were excluded from the 1976 agreement.) The new rector Carl Rudolf Pfaltz, himself a professor of medicine, wrote earlier this month in apocalyptic terms about the future of the medical faculty if some broader base for the university clinic is not found.

But professors in other faculties (such as those at the Biozentrum) and even medical students do not feel threatened by the dispute. After all, research money comes primarily from outside, for example from the Swiss National Science Foundation.

The university's small size — it had 6,712 students in winter 1987–88 — could be a limitation, but it is mitigated for faculty members by a cosy relationship with administrators, says astronomer Gustav Tammann.

Students have recently been granted a wider menu of courses by the Regio co-operative programme, established in 1987 by five universities in the upper Rhine valley — Basle, Freiburg and Karlsruhe in West Germany and Strasbourg and Mulhouse in France.

S.D.

Basle Biozentrum

Still room for the individual

THERE is an old saying in Switzerland: in Zurich, you have to be a banker; in Berne, a general, and in Basle, a professor. Jeff Schatz, himself a professor at the University of Basle Biozentrum (Biocentre), explains that in Basle, people value learning for learning's sake. Professors are regarded highly, but not in the authoritarian tradition of the German university; instead, social status emerges in the form of trust. "If you say you're a professor, they'll lend you money here", says Schatz.

After just 17 years, the Biozentrum has become a part of the academic tradition — and has established a few traditions of its own. An interdisciplinary centre for biological research and teaching, the Biozentrum is a rare bird indeed in German-speaking Europe.

The open administrative structure of the Biozentrum is reflected in its architecture. A central atrium reaches from the ground floor to the top of the eight-storey building, a design that encourages the occupants to meet and talk. The centre was modelled on a Roman villa, says Alfred Pletscher. While director of research for Hoffmann-La Roche in the late 1960s, Pletscher played midwife to the Biozentrum. He and others ensured that the organization was horizontal, with research groups limited to about fifteen people; a rotating directorship; and the opportunity for young researchers to work independently of the senior people. This last sets the Biozentrum apart from many European research establishments.

Although at first its proponents intended the Biozentrum to be primarily a research institution, this proved politically unworkable. Teaching has become an integral part of the programme, and a degree in cellular and molecular biology is an increasingly popular choice among students. Roughly half of Basle's 80 to 100 biology students each year do a 4-year Diplom course at the Biozentrum; the rest follow the traditional plan of zoology, botany and the like. A Diplom is nearly the equivalent of a US master's degree.

The highlight of the programme is a third-year laboratory and lecture course taught sequentially by Biozentrum professors (including industry researchers) on everything from biophysics to immunology. The course has proved a worthwhile investment, says current Biozentrum director Walter Gehring: "I find a lot of my graduate students there."

Research at the Biozentrum is internationally recognized. Best known, perhaps, are Gehring for his work on developmental genetics in *Drosophila* (fruit flies) and Schatz for his work on membrane biogenesis.

Two more bright spots are Werner



Professors Jeff Schatz (top, at morning coffee) and Walter Gehring.

Arber and Eduard Kellenberger. While working in Kellenberger's laboratory in Geneva, Arber was involved in the discovery of restriction enzymes, for which he later shared the Nobel prize. (These enzymes have become a fundamental tool in molecular biology and genetic engineering.) Arber and Kellenberger are among five Swiss out of the 16 group leaders at the Biozentrum.

More recently in the news was physician Urs Meyer, who was one of a group of US and Swiss researchers to identify a previously unrecognized genetic defect in 5–10 per cent of the white US and European population. The defect prevents the body from metabolizing an antihypertension drug because of the absence of intact copies of a liver cytochrome. The discovery has important implications for drug development and cancer research.

The list of other important groups at the Biozentrum would be too long to print.

Biozentrum professors are circumspect about the centre's future despite its impressive track record. "Our main task is to prevent hardening of the arteries", explains Schatz. "There was tremendous vitality when the Biozentrum was set up; researchers came from all over the world." The problem now is that many of the first crew will retire within a few years of each other by the late 1990s. Says Schatz, "We have to break the synchrony beforehand".

S.D.

Basle Institute for Immunology

A paradise on the Rhine. . .

THE organizational principle of the Basle Institute for Immunology (BII) can be found on the back of a French coin: *Liberté, Egalité, Fraternité*. Three Nobel prizes and wide acclaim in the scientific world attest to the wisdom of former director Niels Jerne in setting up the institute the way he did in 1971. Although BII could not exist without the support of Hoffmann-La Roche nor without the hand-picked researchers, it is its organizational structure that sets it apart.

Liberté. Although BII belongs to Hoffmann-La Roche (Roche for short), the researchers are granted "full academic freedom" to publish, teach and collaborate with outside groups of their choice. The support is generous, with an annual budget of SFr24 million for just fifty members, who often use the word "paradise" to describe conditions at BII.

Temporary members (on two- to five-year contracts) must first present an idea that the director considers worth exploring. They are paid well and provided with laboratory space and supplies. Then they are expected to produce. Jean-Claude Weill, hired recently as one of the ten permanent members, said the system was "mounted in a shrewd way by Jerne — there's no excuse not to work". Another



BII boasts this sculpture, "DNA helix" by Jean Tinguely.

member called this realization "the ultimate face-à-face".

The short contracts can lead to too much orthodoxy, says one member, encouraging temporary members to choose projects that are likely to work rather than taking risks. It is one of the few criticisms to be heard about BII.

Despite the initial pressure, the BII structure allows for individual styles, as well as for projects that take longer than two years. Nobel laureate Susumu Tonegawa, who took the 1987 prize for his work at BII on the generation of antibody diversity, was one of the few to make the transition from temporary to permanent member. (Tonegawa left BII for Massachusetts Institute of Technology in 1981.) Members say that the institute's structure and Jerne's support were pre-

requisites for Tonegawa's success.

Egalité Both Jerne and current director Fritz Melchers consider themselves heirs to the spirit of physicist Niels Bohr, who encouraged his students to display a "joyful disregard for authority", says Melchers. "We try to have as little hierarchy as possible" to allow for the most interaction. The high turnover prevents a class system from forming. The only one who stands above the rest is Melchers, who divides his time between his administrative duties and the laboratory.

English is spoken and researchers from all over the world (fewer than 10 per cent are Swiss) use exclusively first names. Outside work time (which spans the entire day and most of the night), the institute plays together too. Not long ago, the staff produced an opera on the immune system in which nearly everyone participated. Even so, original members such as Charles Steinberg (who came to BII from Oak Ridge National Laboratory in Tennessee) say the atmosphere now is more subdued than at the start, "when you would come in the morning and find someone sitting in a tree outside the institute".

Fraternité. The 'unit size' of a research group at BII is two people: one member and one technician. This forces collaboration among groups, which can be both a blessing and a curse. According to member Louis Du Pasquier, "The path has twists and turns . . . the turnover of people is high and in long-term projects, permanent members cannot always follow their own line as they would with a stable team." Melchers replies: "the institute profits a lot, even if the individual gets a bit frustrated". In terms of productivity, Jerne's grand experiment has been a success. Jerne himself won a Nobel prize in 1984 for his theories concerning the specificity in development and control of the immune system; he is recognized as the spiritual father of modern cellular immunology. He shared the prize with Georges Köhler and César Milstein. Köhler completed his doctoral dissertation at BII in 1974 under Fritz Melchers before going to Milstein's group at the MRC Laboratory of Molecular Biology at Cambridge (UK), where they succeeded in creating the first monoclonal antibodies. BII continues to "develop hidden talents", says Melchers, and turn them loose on the scientific world.

But what's in it for Roche? Former Roche research director Alfred Pletscher cites three goals that Roche owners and management had in mind when BII was conceived in 1968. First came "ethics": Roche was number one in the world in pharmaceuticals and "wanted to give something back" to science. Second, there

Tennis, anyone?

THERE have been just two directors at BII, Niels Jerne (who left in 1980) and Fritz Melchers. Jerne, a Dane, came to BII from the Paul Ehrlich Institute in West Germany, where he had been frustrated by West German bureaucracy. He created the institute with as little bureaucracy as possible, convinced that this was one key to success in research.

Jerne's management style was one of single-minded intensity. Members compare him to a "golf player"; he pursued a few problems in a linear fashion, such as the generation of antibody diversity, while ignoring others. "He left me with a sense of awe", said Louis Du Pasquier, who has been there since the beginning. "He set a standard so high, it was almost frightening." Says Alfred Pletscher, some people at Roche were "not enthusiastic" about hiring him, calling him "a philosopher". But in retrospect, Jerne's record confirmed that he was the perfect choice.

Melchers, a West German, is more of a "tennis player". A pragmatist who has brought more diversity to the science at BII through his appointments, Melchers relies on continual interaction with members to plot his strategy. More approachable than Jerne, "he lets you do your work and trusts you to do it well", says Du Pasquier. Melchers explains that he wants to keep many lines of inquiry open, despite the fact that most young researchers care only about the genetic approach.

"There are not enough cellular immunologists left; everyone wants to clone a gene. My job is to balance their desire to jump on this bandwagon against long-term interests I think we need to pursue." S.D.

was the thought that BII could become an important source of talent for the corporate research division of Roche. And finally, there might be patents from which Roche could profit.

In practice, Roche has gained most by hiring BII alumni and using BII as an "open window", as Melchers puts it, onto modern biomedical science. Contact with BII ensures that Roche corporate research does not become "bureaucratized". As a bonus, BII provides the odd patent.

The worst fear among BII members is that someone might turn out the lights in paradise. Melchers acknowledges that the institute's future is dependent on the goodwill of the owners of Roche, who have protected BII before when management had threatened to close it. Melchers admits that there is a chance BII would be disbanded "if we are no longer a centre of excellence in our field". But the lack of a guarantee from Roche also provides an incentive, which, he says, is not all bad. After all, with a permanent lease on paradise, who would work? S.D.

Lausanne polytechnic

Latecomer exudes practicality

SWITZERLAND's *other* polytechnic speaks French and is at Lausanne, bordering Lake Geneva (Lac Léman). It should not be confused with the University of Lausanne — both campuses lie alongside the same coastal highway.

Compared with the ETH at Zurich, Lausanne is a newcomer. Indeed, it owes its existence (since 1969) to the confederation's wish to expand its role in higher education beyond the constraints of the Swiss constitution, which gave it the right to run only a single university.

In 1969, the confederation took over from the University of Lausanne the polytechnic institute which had, itself, begun life as a private engineering institution — and then found its plans frustrated by defeat in a referendum, largely in the face of the argument that all members of a university should have some say in its government. So both polytechnics are now supervised by a Council for the Federal Polytechnics (see below), under a full-time president and with the president of each university a vice-president.

Even more than its German-speaking twin at Zurich, the Lausanne polytechnic exudes practicality. It offers its potential engineers and architects some instruction in law and in problems of the environment and of development in the third world, but

Zurich teaches humanities and languages to all engineers. From its recent start, the Lausanne polytechnic has grown to 3,200 students, or to less than a third the size of Zurich.

As in many other places, the traditional branches of engineering are more or less static, but the past few years have seen a boom in applications for places in physics (contrary to experience elsewhere) as well as *informatique* (computer science), begun as recently as 1981, and Lausanne's own special interest — '*microtechnique*', which encompasses not only the manipulation of semiconductors, but the design and development of robots (see page xxx).

One unexpected feature of engineering diplomas from the two Swiss polytechnics is that they are professional qualifications in their own right, without the usual validation of courses by professional organizations and post-certification apprenticeship. The government has for the past decade taken power to regulate professional standards, using its unique position as the supervisor of the institutions at which most engineers are trained as a guarantor of standards.

After a period of uncertainty occasioned by the low birth-rate between 1965 and 1975, the polytechnic now seems assured again of continued growth. There will be a

further 20 per cent increase of student numbers in the next four years, but the institution's managers do not believe that will be the end. Foreign students are a large proportion of the total.

Among the students from elsewhere is a group of 10 or so from the Carnegie-Mellon Institute at Pittsburgh in the United States, with which Lausanne has an exchange programme. Students make the swap in the third year of their regular courses, which usually occupy them for 4.5 years, and are said to profit sufficiently from the experience to justify the continued spending of SFr75,000 a year.

While the practical atmosphere at Lausanne is chiefly manifested by the research programme, students are encouraged to lend a hand both in the polytechnic's own laboratory work and as entrepreneurs in their own right. For the past five years, the university has encouraged its students to take on 'junior enterprise contracts' with outside organizations, usually companies, using the institution's laboratories and equipment to earn contract sums of between SFr2,000 and SFr10,000.

The practical involvement of the polytechnic at Lausanne is partly determined by the generosity of its staffing. While the number of professors is only 120, there are more than 600 other qualified people on the polytechnic's own books and a further 500 supported by funds from other sources.

Another reason for Lausanne's engagement in practical matters is the ethos of the institution. If its chief product, its stream of graduates, is meant to benefit Swiss industry, why should not the same establishment be used as a means of providing the strategic research on which the health of industry depends? The obvious constraint is that federal funds cannot be used for commercial purposes, but there is nothing to prevent federally supported institutions from signing research and development contracts with industry.

The difference between the two polytechnics is best illustrated by what they each say about their conduct of applied research. Zurich is proud that innovations arise "from the bottom up" (*de bas en haut*), but Lausanne is equally attached to the notion that departments take the initiative. To help them, there is a liaison committee including representatives of Swiss companies, together with an applied research and development company which has reached three major agreements with industrial consortia to carry out applied research in fields such as the use of lasers in materials processing and in the use of opto-electronics for the processing of medical diagnostic images.

The result, in which the Council for the Federal Polytechnics clearly concurs, is a decision that a substantial part of Lausanne's income should be spent on

Council for the Federal Polytechnics

THE Council for the Federal Polytechnics, the umbrella for the polytechnics at Zurich and Lausanne, is now much more than that. Over the years, it has also accreted five substantial basic research laboratories, which the council manages. This seems to be an arrangement unique to Switzerland; it is as if, in the United States, the national laboratories were supervised by an independent agency responsible only to the cabinet which also supervised MIT and Carnegie-Mellon.

The common justification of the arrangement is that it allows the confederation to follow flexible policies in the support of important or emerging fields, but it is historically an accident. More than a century ago, ETH Zurich had assumed responsibility for the Swiss Materials Testing and Research Laboratories and the Forestry Research Institute.

When, after the Second World War, it seemed prudent to set up reactor research and nuclear research institutes (in 1955 and 1968 respectively), there was no other structure in which to fit them. Since then (in 1977), the council has also taken under its wing the Federal Institute for Water Resources and Water Pollution Control,

has controversially decreed the merging of reactor research and nuclear physics into a single Paul Scherrer Institute and has singled out the plasma physics institute at Lausanne for care and custody.

More recently, the council's affairs have not been trouble-free. In 1984, it commissioned from the Swiss industrial consultants Hayek a report on its own efficiency and effectiveness, winning the two-edged response that it should be possible to save between 250 and 300 posts in both the administration of its dependants' affairs and in teaching therein, but that the proper execution of its institutions' functions would require the recruitment of more than 1,000 people somewhere in the system.

The federal council seems to have accepted the essence of the Hayek report's recommendations whence, in part, the continuing resumed growth of the two polytechnics, costing the confederation this year SFr598 million. The associated research institutes will similarly cost SFr197 million this year, of which all but SFr44 million will go on research and development, and may earn up to SFr40 million for testing fees and the like. By Swiss standards, this is big business. J.M.

research. The calculation is that the total value of research at Lausanne in 1987 was SFr124.4 million, of which SFr47 million came from external sources and the remainder from the regular budget.

Lausanne has also in its short lifespan accumulated a group of extra-departmental functions that occupy a third of the total staff on site, and which range from the plasma physics laboratory (page 328) to a land management group. A year ago, it was boasting of having become one of the centres in the international network of a dozen institutions collaborating on the numerical solution of problems in fluid mechanics. The polytechnic, the proud possessor of Cray and Cyber supercomputers, is also expected to provide supercomputer facilities for other Swiss universities, but manages itself to use 38 per cent of the available computer-time.

In the circumstances, it is odd that the polytechnic seems to have shared the general Swiss failing to appreciate the importance of the computer revolution. That this is so is nevertheless openly acknowledged. The student course in *informatique* was begun in 1981, and is already the third most popular.

Implicitly, this is a criticism of the system of management of the two institutions at Lausanne and Zurich, although the general opinion is that Swiss polytechnics would have awakened sooner to the need for trained computer people if Swiss industry had been quicker off the mark. "Maybe we were too slow", says Dr G. A. Grin, special adviser to the Council for the Federal Polytechnics.

Meanwhile, Grin acknowledges a different need for change. With the customary deliberation of the Swiss legislative process, proposals for reform of the council are already being canvassed. Originally intended as a means by which the confederation could support two universities without amending the constitution, the council has formally to approve the creation of new posts and even individual appointments to existing senior posts. New courses have also to be approved by the council. Similarly, sitting as it does directly on the confederation's organizational chart, the council must enforce the usual civil service rule that money unspent in one financial year should not be rolled over to the next.

On the face of things, all this suggests that the Swiss polytechnics, and that at Lausanne in particular, are but devices for bending academics' interests to the needs of industry. That they (and their institutes) perform such services is not in doubt, but there is also evident respect for the doctrine of "freedom in teaching and research" repeated in all legislation on the subject, not least in the research law of 1984. If Lausanne were better known, there would be many governments seeking to imitate it.

J.M.

Friedrich Miescher Institute

Plans for rejuvenation

JUST one hundred years after Friedrich Miescher discovered nucleic acids in Basle in 1869, the two independent pharmaceutical producers CIBA and Geigy agreed to found a basic research institute named after him. Although the research at the Friedrich Miescher Institute (FMI) has not quite lived up to that of Miescher himself, FMI has played a large role in training Switzerland's next generation of biologists.

FMI is strongest in plant sciences. Molecular biologists Barbara and Thomas Hohn started to work on plants when they joined FMI in the late 1970s. They had achieved distinction in discovering that bacteriophage lambda DNA is packaged into preformed virus heads, a discovery which led to the development of 'cosmids' by Barbara Hohn and John Collins. Barbara Hohn has shifted her focus to genetic flux in plants; her husband Thomas has worked lately on molecular plant virology using cauliflower mosaic virus. Fred Meins is known for his studies of cell commitment in plants.

New director Max Burger wants to rejuvenate and 'focus' FMI. A Swiss national, Burger comes to the job with 12 years of experience in the United States and 15 years at the Basle Biozentrum. His own work is in cell-cell interactions and cancer. He sees his main tasks as re-establishing neurobiology at FMI and hiring young people in areas he considers "top-

heavy". Burger would like to increase the number of "junior groups" from 20 to 40 per cent of the total.

Another task Burger has set himself is making relations with the university even better. There are more than 50 doctoral candidates at FMI and a couple of dozen more students at earlier stages — far more than at, say, the Basle Institute of Immunology. The students, most of whom are Swiss nationals come into contact with postdoctoral fellows from a wide variety of countries. The group leaders are quite international, too — just 5 out of 22 are Swiss.

If there is a 'sword of Damocles' hanging over FMI, it is the spectre of immigration restrictions. So far, explains Thomas Hohn, himself an Austrian, this has been a problem only in the area of student visas, where abuse of the privilege has led to tighter controls. But xenophobia in Basle would not come as a great surprise; the foreign population has risen to around 30 per cent (the average in Switzerland is 15 per cent).

Scientific relations between FMI and the now unified Ciba-Geigy are good but "could always be better" on both sides, says Burger.

However, the "burning issue" is the location of the institute. Despite the company's assurances that a new building can be found, FMI continues to be located inside the security gates of Ciba-Geigy. S.D.

Industry/university cooperation

Room for further improvement

EVEN in Switzerland, researchers express discontent from time to time. Despite the appearance of being a paradise for researchers, with solid funding, little grant-writing and openness to foreign researchers, Switzerland does not necessarily seem so idyllic to its inhabitants. "Not as nice as Singapore" cautions one biologist; "no room for Swiss researchers returning from abroad" whisper others; "times were tougher eight years ago; who knows about the future?" admonish funding-war veterans.

But, perhaps significantly, nobody denies the benefits of industry support for basic research. The Swiss chemical industry has poured millions into university research with no strings attached and universities and industry alike have reaped the rewards.

The organic chemistry department at ETH in Zurich shines brightest in this regard (ETH, the Federal Institute of Technology, page 332). ETH professor Tadeus Reichstein performed the first chemical synthesis of vitamin C in 1933

with the help of industrial funds, a procedure that was scaled up by Hoffmann-La Roche and brought the company tremendous profits.

1939 Nobel laureate Leopold Ruzicka had support from Ciba and other companies when he did his prize-winning work on isoprenes in the early part of the century. Present-day ETH chemistry professors continue to draw support for their widely respected work from the Basel chemical industry.

Former department chairman Oskar Jeger explains that success was not a requirement to receive support. "Industry's interest was in supporting the best possible education to provide itself with manpower. It was always extremely liberal, with no strings at all." The Basle chemical companies — Hoffmann-La Roche, Ciba-Geigy and Sandoz — continue to support research in chemistry, not only at ETH but also at other universities. Ciba-Geigy, for example, supports 300–400 graduate students, half in chemistry and the rest in economics, social sciences,

medicine and biology.

Apart from ETH, Basle is the university that has profited the most from the Swiss chemical industry. The benefits are most visible at the Biozentrum, which was built with the help of a 'catalytic unit' of SFr5 million according to Alfred Pletscher, who helped to create it.

There is continuing interaction in Basle in the form of industrial researchers who are also professors and of graduate students who do their thesis at a company. Basle Institute for Immunology members "do all the immunology teaching" in the natural sciences faculty at the University of Basle, says institute director Fritz Melchers. Sources on both sides say that the students are not "abused" by their advisers and that industry laboratories remain open to the university.

Jakob Nüesch, Ciba-Geigy's pharmaceutical research director, says that strong contacts between industry and academic institutions are "part of our tradition. New trends in industry must have their roots in university developments". Ciba-Geigy's pharmaceutical research department collaborates with eight groups in Basle as well as with researchers in Berne, Geneva, Zurich, Lausanne and with groups "from Texas to Canada", seeking "projects of mutual interest," says Nüesch. "We don't just place orders." About a third of the graduate students and postdocs at Ciba-Geigy return to work for the company.

When asked who gains more from the cooperation of industry with universities, the Biozentrum's Jeff Schatz calls it "a tie". Industry contributes to a "critical mass" of biomedical researchers and to the international atmosphere in Basle research, and industrial efficiency rubs off on university government. In turn, the three chemical giants have a steady supply of talent.

In other areas, university officials admit that the interaction between universities and Swiss industry could stand improvement. ETH, the University of Basle and others have established 'contact offices' which provide a list of past, current and future research projects and matches companies with research groups for collaboration.

Ralf Hütter, ETH vice-president for research, points to the need to improve contacts in electronics, for example. Some ETH institutes are at present supported with up to a quarter of their budget by industry, but others have no support at all.

Hütter would like to see a general industry support level of 15–20 per cent but no more. In the past, too much industry money has been put into short-term projects meant to yield quick solutions.

On the university side, Hütter says "we must educate people not to fear industry". He sees his task as "a challenge to establish a tradition in those fields where there was none before".

S.D.

Engineering

Living in a material world

LAST year's merger of two electrical engineering giants into Asea Brown Boveri (ABB) has breathed life into the Swiss half of the company. Although some rough spots lie ahead as 'restructuring' proceeds, Maurice Campagna, the new research director, foresees an improvement in ABB's interaction with Swiss universities, giving benefits on both sides.

The merger brought together Brown Boveri & Cie, a Swiss machine and electrical systems company nearly a century old, and its Swedish rival ASEA. The resulting hybrid is one of the leading European providers of gas turbines and nuclear and conventional power plants as well as transportation and communication technology. Among other activities, it has helped to build a prototype underground train now being tested by the London 'tube'.

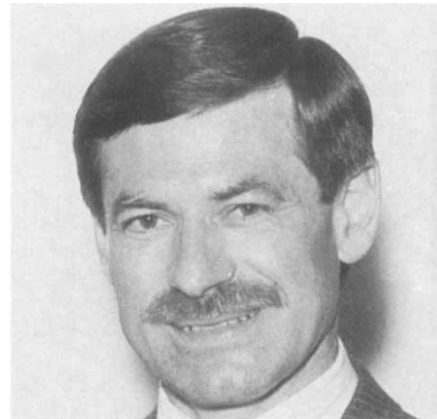
ABB is the self-proclaimed 'flagship' of the Swiss machine and metal industry, investing as much as SFr1,000 million a year in research and development. Companies producing machines, metals and electrical systems invest nearly half of the money put into research and development by Swiss industry each year (most of the rest comes from the chemical industry).

ABB research in Switzerland revolves around a centre in Dättwil, 30 km northwest of Zurich, which features research into environmental technologies, materials science, semiconductors and electronics. The location, says Campagna, makes it easy to attract foreign researchers — some 60 per cent of the scientific staff — who are then "willing to work harder" because of the close proximity of the Alps.

The 42-year-old Campagna, a Swiss national from the Italian-speaking canton of Ticino, joined ABB after a brief tenure as a professor of physics at ETH. Like so many ambitious Swiss, Campagna had left

Switzerland after his PhD for seasoning abroad. He returned in 1986 after five years at Bell Laboratories in the United States and nine years at the Jülich Nuclear Research Establishment in West Germany.

A key task for ABB, he says, is to attract graduates away from high-glamour fields like biotechnology or pharmaceutical development by providing projects both "scientifically interesting and close to

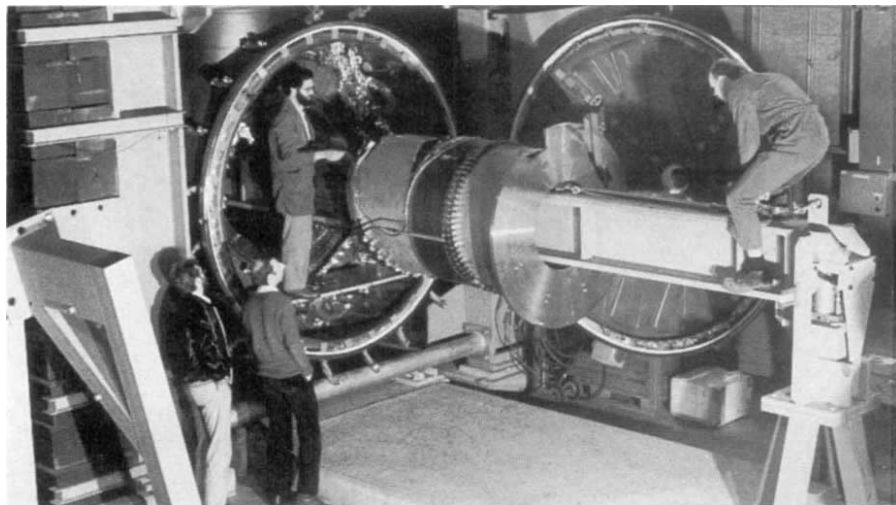


Maurice Campagna: "even ETH" could pay more attention to corporate needs.

the market". The growing share of ABB research into areas such as environmental engineering, in addition to the traditional areas such as gas turbines or power plant design, might provide the necessary pull.

At the same time, Campagna sees a need for researchers at the universities, even vaunted ETH, to pay more attention to corporate needs. Breakthroughs at ETH in semiconductor technology, for example, were not followed up (although the lack of a Swiss semiconductor industry also had something to do with this). For research in electrical engineering, Campagna says, ETH has been a "black hole."

S.D.



High-field test apparatus made by ABB for the Swiss Paul Scherrer Institute for the testing of superconducting cables.

Ciba-Geigy

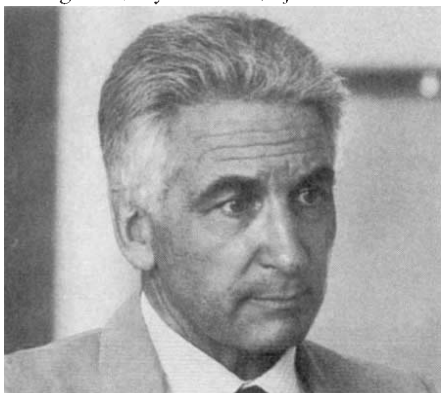
Synergy of chemists and biologists

THE introduction of genetic engineering techniques has turned pharmaceutical research on its head, and Ciba-Geigy's Jakob Nüesch has seen the effects at first hand. At first, there was antagonism between chemists — traditionally the top dogs in drug development — and molecular biologists, whose work came as "a shock", says Nüesch, to a rigidly organized industry.

At first, says Nüesch, director of Ciba-Geigy's pharmaceutical research division, pharmacologists reacted with a disbelieving "Oh, come on!" when told that custom peptide biosynthesis could now be done in seconds by genetically modified micro-organisms. They wagered that human insulin, for example, would never be made by these methods.

The success of 'genetic engineering' is no longer in doubt, and the "tension" has given way, says Nüesch, to "an excellent integration" between chemists and biologists that is "a joy to watch". Chemists have found a "new self-confidence" after the crisis and both sides have recognized the need for cooperation.

"The chemists have to become more biological", says Nüesch, "just as the biol-



Nüesch: presiding over a time of rapid change. ogists have become more chemical". Through it all, Ciba-Geigy has maintained its position as one of the top three producers worldwide of dyestuffs, pharmaceuticals and agricultural chemicals. Nearly 3,000 of the 11,000 employees in Basle are involved in research and development. The company was formed in 1970 by the merger of CIBA, where epoxy resins were first developed, and Geigy, whose chemist Paul Müller had discovered the insecticidal properties of DDT, for which he received the 1948 Nobel prize for Physiology or Medicine. Overall, there are 24,000 employees in Switzerland and 82,000 in the world.

Although half the university graduates hired by Ciba-Geigy are chemists, there are more postdoctoral fellows (whose work is nearer the cutting edge of basic research) in the biological sciences than in

chemistry. This has to do with the different "maturities" of the two fields. But despite the increasing investment in biotechnology, controversial treatments like somatic gene therapies are "a really long-term thing", says Nüesch. There has been unwarranted optimism in this area, unjustified because "so few diseases are the result of a single gene". Nüesch rejects the idea of experimenting on the human germ-line, saying "even if we saw a possibility to develop therapies in this direction, we wouldn't do it".

By contrast, a major growth area in pharmaceutical development is neuro-molecular biology. This comes as no surprise at Ciba-Geigy, which has traditionally been a major producer of psychoactive drugs such as tranquillizers and compounds against epilepsy. S.D.

Hoffmann-La Roche

Biotechnology now to the fore

INTERNATIONAL pharmaceutical giant F. Hoffmann-La Roche & Co. (Roche for short) invests a fifth of its pharmaceutical division budget in research, more than its Basle brethren Ciba-Geigy and Sandoz. The company is moving rapidly into biotechnology and genetic engineering despite growing public disquiet about the consequences (see page 330).

With more than 6,500 employees at its corporate headquarters and research and development and production facilities in Basle (out of 46,000 worldwide), Roche must be concerned about the views of the citizens. A series of public forums has been held on subjects such as the side-effects of the acne medicine Accutane, and the effects of degenerative neurological diseases. The forums feature top Roche executives, says spokesman Eckart Gwinner, and are meant to "show how we can unify our commercial goals with the health of the population".

Roche made its name as the first large-scale producer of vitamin C in the 1930s and vitamin A in the 1940s. 'Vitamins and fine chemicals' still account for more than a quarter of the total sales. The company also derived big profits from sulphonamide antibiotics beginning in 1949 and from the benzodiazepene tranquillizers that are the active ingredients in Librium (introduced in 1960) and Valium (1963).

An investment in biotechnology began to pay dividends in 1986 with the introduction of interferon alpha-2a, licensed under the name Roferon-A. The compound has been licensed in several countries as a treatment of hairy-cell leukaemia, Kaposi's sarcoma and other cancers.

But modern molecular biology at



Ciba-Geigy dominates the Basle waterfront, employing 11,000 people in the city.

Roche did not really take off until Hermann Bujard was hired to lead the biology department. Bujard, currently director of the Centre for Molecular Biology in Heidelberg, West Germany, assembled an excellent research staff in the early 1980s. Publication began to be encouraged, and interaction with the Basle Institute for Immunology (BII) was promoted.

The present head of biology research Michael Steinmetz, himself from BII, explains where Roche is headed: "The



Alfred Pletscher, director of research at Roche in the 1960s when the Basle Biozentrum was established (page 334). Now president of the Swiss Academy of Medical Sciences.

trend now is to find novel chemical compounds that replace or inhibit factors that play a role in immune response, such as interleukins and interferons." For example, Roche researchers are looking for a non-peptide interferon-gamma antagonist in microorganisms.

Other projects under way in the biology department include the search for a malaria vaccine (which even if found will probably not recover the investment made) and development of AIDS therapeutics and diagnostics. S.D.

High-altitude research

Cold comfort on Jungfrauoch

THE multifaceted research station located at the top of the Jungfrauoch might not be there but for tourism. In 1896, industrious Swiss engineers completed the last leg of a cog railway that runs inside the notorious Eiger on its way to the "highest railway station in Europe" at 3,550 m.

Jungfrauoch (the name means 'yoke of the virgin') has been one of Switzerland's most popular tourist attractions ever since. The three-hour trip from Interlaken (involving two changes of train and ever narrower tracks) is accompanied by sing-song announcements in six languages (including Japanese but not Swiss-German) telling visitors what they can expect to see. At the top, people queue for an hour to ride in the lift to the Sphinx observatory, which sits on a rocky outcropping overlooking the Aletsch glacier and La Place de la Concorde Suisse.

Only after one climbs the stairs into the cluttered eyrie of the solar and atmospheric physics group of Luc Delbouille and Ginnette Roland of the University of Liège, Belgium, is the babble of tourism replaced by the earnestness of research.

Delbouille and Roland have worked here since 1958, when the large solar spectrograph was built under the direction of then-group leader Marcel Migeotte. Migeotte's 1951 observations at Jungfrauoch proved the presence of methane in the atmosphere. The amount of methane has been increasing since at a rate of 0.7 per cent a year, although the significance of this result is still unclear.

Jungfrauoch turns out to be an ideal location for solar and atmospheric physics not only because it is high up — above one third of the Earth's atmosphere — and far from sources of air pollution, but also because the air is consistently drier here than anywhere else on Earth except on the Antarctic plain. This is important because water distorts the spectral measurements.

Although Delbouille *et al.* came to Jungfrauoch to do solar physics, a lack of

support has caused them to shift their focus to the Earth's atmosphere. Delbouille is not happy about the trend, which has also been manifested in the closing of the Sacramento Peak Solar Tower in the United States. The Jungfrauoch groups have been known for their series of atlases of solar spectra at a variety of wavelengths, beginning with an atlas of infrared spectra by Migeotte.

Observing the atmosphere has turned out to be politically volatile. Jungfrauoch has been a key centre for gathering data on the presence and effects of chlorofluorocarbons (CFCs) on the stratospheric ozone layer. One of the most striking observations is that CF_2Cl_2 , or CFC-12, was not seen at all in 1950 spectra but is seen as a visible peak now. One CFC breakdown product, hydrofluoric acid or HF, is increasing in abundance by 8.5 per cent a year according to data gathered at Jungfrauoch by Rudy Zander. Unpublished data show that the trend continues beyond 1986, the last year shown in the figure below. The environmental significance of the data remains in dispute, especially because of large seasonal and daily fluctuations.

Jungfrauoch is also an ideal place to observe so-called 'greenhouse gases' like carbon dioxide, which is produced by the burning of fossil fuels as well as by natural processes. These gases are thought to contribute to the warming of the Earth's surface and eventually, researchers fear, to global climate change. But despite a unique record of measurements dating back to 1951, the Jungfrauoch group has been unsuccessful in garnering European Community support for the tracking of greenhouse gases. "Money is tight" in the EC, says Delbouille.

Should another source of support become available, the solar observatory could double its observing season by including the winter months, a move Delbouille would welcome.



The Sphinx observatory overlooks the snowy expanse of La Place de la Concorde Suisse.

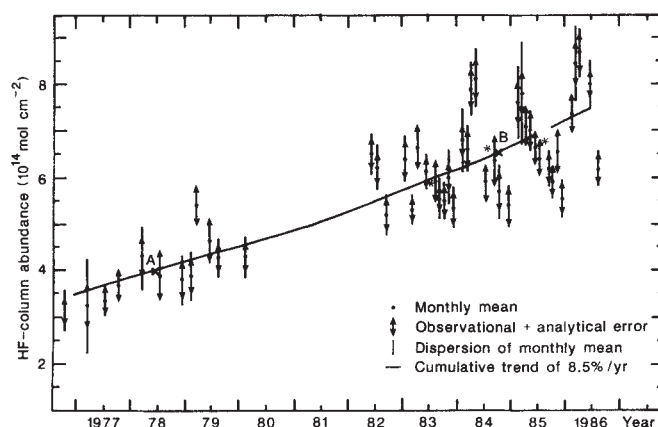
The other projects at the High Altitude Research Centre run the gamut from astronomy to physiology. Perhaps best known is a group from the University of Berne led by Hermann Debrunner which studies cosmic rays: high-energy particles from space, mostly protons, which collide with the atmosphere, sending down showers of secondary particles. Debrunner and his colleagues are interested in detecting the occasional particles originating in the Sun, as distinct from the background particles from the Galaxy and beyond. Debrunner's group was the first to detect high-energy neutrons in solar radiation using Earth-based monitors in 1982, a discovery that was confirmed by satellite observations.

The study of cosmic rays at Jungfrauoch dates back to the early part of this century. A group led by P. M. S. Blackett from the University of Manchester in 1950 installed a 14-tonne electromagnet and a cloud chamber to detect charged particles from the cosmos.

Debrunner's equipment is sufficiently automated that only 17 person-days had to be spent at the Research Station last year. Normally, the data flow directly into a microcomputer and are sent to Berne by telex.

Astronomers from the University of Geneva have used Jungfrauoch and other observatories to measure the light from 28,000 stars through seven coloured filters. They have created a data bank that has become an international standard source. The Geneva group continues to work on determining the parallax of stars (the apparent "movement" of an object when it is viewed from two or more points) to determine their exact distance. These stars could then be used as reference points in the sky to determine the distance of other heavenly objects.

S.D.



Hydrofluoric acid measurements from Jungfrauoch, showing a steady increase since 1977. (Data from R. Zander *et al.*, *J. Atmos. Chem.* 5, 385-394; 1987.)

Magnetism of spiral galaxies

Alexander Ruzmaikin*, Dmitry Sokoloff† & Anvar Shukurov‡

* Institute of Terrestrial Magnetism, Ionosphere and Radiowave Propagation, Academy of Sciences, Troitsk 142092, Moscow Region, USSR

† Physics Department, Moscow State University, Lenin Hills, Moscow 119899, USSR

‡ Space Research Institute, Academy of Sciences, Profsoyuznaya ul. 84/32, Moscow 117810, USSR

The concept of the turbulent dynamo in interstellar gas combined with the discovery of global magnetic structures in spiral galaxies has led to a consistent picture of galactic magnetism. Large-scale magnetic fields are generated and maintained by helical turbulent motions of interstellar gas and by differential galactic rotation. Seed fields for the dynamo can be produced by outflows from supernovae and hot young stars.

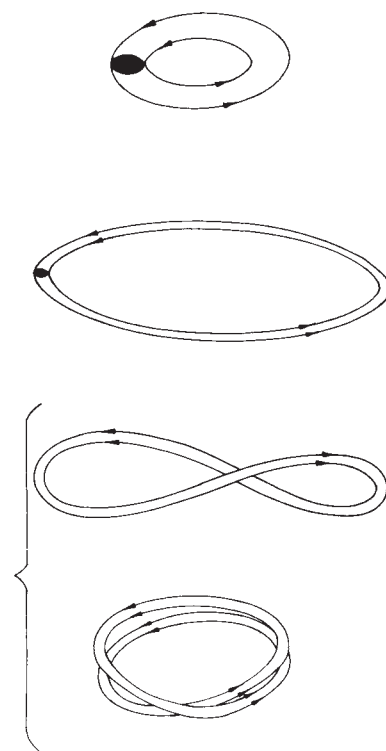
GALAXIES are the largest magnets in the Universe, with scales naturally measured in kiloparsecs ($1 \text{ kpc} = 3 \times 10^{21} \text{ cm}$). Because we are living within a spiral galaxy, the Milky Way, investigation of the magnetic fields of spiral galaxies has taken a special place in the study of cosmic magnetism, but magnetic fields are a universal property of all galactic-type objects, as is evident from the ubiquity of synchrotron emission from relativistic electrons gyrating in magnetic fields. The past ten years have been notable for rapid, qualitative progress in understanding the magnetism of spiral galaxies, a result of both theoretical and observational developments. A few decades ago, galaxies were considered as combinations of stars; forty years ago the role of interstellar gas and cosmic rays was realized, but only now are magnetic fields considered an equal component of the interstellar medium.

To a solid-state physicist, for example, the strength of galactic magnetic fields would seem miniscule. Although interstellar magnetic fields are as weak as few microgauss, in rarefied interstellar gas their energy density, $H^2/8\pi \approx 10^{-12} \text{ erg cm}^{-3} \approx 1 \text{ eV cm}^{-3}$, is comparable to the thermal energy of the interstellar medium $nkT \approx 10^{-12} \text{ erg cm}^{-3}$ ($n \approx 1 \text{ cm}^{-3}$ and $T \approx 10^4 \text{ K}$), the energy density of interstellar turbulence $nm_p v^2/2 \approx 10^{-12} \text{ erg cm}^{-3}$ (the r.m.s. turbulent velocity $v \approx 10 \text{ km s}^{-1}$) and the cosmic ray energy density $\sim 10^{-12} \text{ erg cm}^{-3}$.

Cosmic magnetism differs drastically in nature from the ferromagnetism of a permanent magnet or from magnetism produced by electromechanical induction devices. The magnetic fields of the majority of astrophysical objects are the self-excited offspring of hydrodynamic motions of an electrically neutral conductive medium, or plasma. The magnetic fields of the Earth, the Sun and spiral galaxies have a common source in the dynamo action of internal fluid motions. Galactic magnetic fields were discovered only recently, but our understanding of their properties is more advanced than that of geomagnetism or stellar magnetism. The reason for this is two-fold. First, we reside within a dynamo region in the Milky Way and are able to observe, at least in principle, all components of the magnetic field. (Note that electrodynamic vacuum boundary conditions forbid penetration of the probably dominant azimuthal axisymmetric components of magnetic fields outside the liquid core of the Earth, the convection zone of the Sun, or the ionized disk of a spiral galaxy.) Second, spiral galaxies are transparent to radiowaves and partly transparent to optical emission. This makes it possible to investigate both hydrodynamic and magnetic properties of external galaxies and of distant parts of the Milky Way.

The formation of regular magnetic structures in turbulent interstellar gas is a specific example of order emerging from chaos¹. It is interesting to note that this order arises in a system that is linear in the magnetic field. The dynamo that creates the large-scale magnetic field acts by violation of the reflection

Fig. 1 This schematic of Zel'dovich's rope dynamo illustrates how a magnetic field can be generated by a helical turbulent hydromagnetic dynamo. A loop of magnetic flux is stretched to twice its radius, twisted and doubled upon itself. Solid shading shows the loop cross-section, which is halved after stretching. After every cycle shown here the magnetic flux of the loop is doubled. If the process can repeat, the flux density can grow exponentially.



symmetry of the turbulence in rotating galaxies; this violation is associated, in turn, with dominance of, for example, right-handed helical motions in the turbulent flow. The generation of magnetic fields is a threshold phenomenon; the analogue of the bifurcation parameter is the dynamo number, proportional to the intensity of the mean helicity of the flow.

The existence of an intrinsic galactic magnetic field was anticipated by Fermi² on the basis of studies of cosmic ray confinement^{3,4}. Observational evidence of galactic magnetism was provided by the discovery of the polarization of starlight^{5,6}, the interpretation of non-thermal (synchrotron) galactic radioemission⁷⁻¹⁰ and the discovery of its polarization¹¹⁻¹³. Comprehensive discussions of observations of magnetic fields in our Galaxy are given in refs 14-16. The most reliable and detailed quantitative information on galactic magnetic fields currently comes from observations of the Faraday rotation of the polarization plane of radio emission passing through interstellar space.

Here we present a fundamental theoretical approach to magnetic fields of spiral galaxies, in which the hydromagnetic dynamo concept plays a key role. Our views on this subject

have been profoundly influenced by discussions with Yakov Borisovich Zel'dovich. The dynamo generation process can be illustrated by Zel'dovich's rope dynamo model (Fig. 1).

Magnetic fields in spiral galaxies

Observation of cosmic magnetic fields is one of the most difficult problems in astronomy, and it is therefore understandable that observational data on magnetic fields in the Milky Way and other galaxies are often ambiguous, uncertain and even contradictory. Nevertheless, some conclusions can be drawn, more or less confidently, from existing observational data.

The Faraday effect is very useful in studies of galactic magnetism because it allows one to determine not only the strength but also the direction of a magnetic field. When an electromagnetic wave propagates through a magnetized interstellar thermal plasma the polarization plane rotates over an angle

$$\varphi - \varphi_0 = 0.81 \lambda^2 \int_0^L n_e \mathbf{H} \, d\mathbf{l} \equiv \text{RM} \lambda^2,$$

where φ is the positional angle of the polarization plane, φ_0 is its intrinsic value within a source (pulsar or radio galaxy) whose distance to the observer is L (pc), λ (m) is the wavelength, n_e (cm^{-3}) is the number density of thermal electrons, $\mathbf{H}$ (μG) is the magnetic field strength and integration is carried out along the path length. The integration factor is called the rotation measure, RM. Under galactic conditions RM acquires measurable values in the radio range $\lambda \approx 0.1\text{--}1$ m.

Statistical analysis of Faraday rotation measures of pulsars and extragalactic radio sources has yielded detailed information about magnetic fields within ~ 4 kpc of the Sun^{17,18}. The large-scale field $\mathbf{B}$ (that is, the field whose scale exceeds the energy-range scale of interstellar turbulence, ~ 100 pc) is dominated by the azimuthal component B_ϕ . In the local spiral arm B_ϕ is parallel to the direction of Galactic rotation. There is ample evidence to suggest that the azimuthal component of the large-scale field has the same direction both above and below the Galactic plane: $B_\phi(z) = B_\phi(-z)$, where z is the vertical coordinate orthogonal to the Galactic plane. The strength of the large-scale magnetic field in the solar vicinity is $B \approx 2\text{--}3 \mu\text{G}$ and the direction of this field, as determined from analysis of Faraday rotations of extragalactic sources, is toward Galactic longitude $106^\circ \pm 8^\circ$ (ref. 17, p. 250); recent pulsar data give instead longitude 96° (ref. 19). This means that the ratio B_r/B_ϕ is 0.1–0.3; the field is almost in alignment with the local spiral arm. Such alignment is typical of spiral galaxies^{20,21}. Information on the vertical field B_z is obscured by statistical noise, except in central parts of galaxies, where B_z is anomalously strong and comparable to the other field components^{22,23}.

Analysis of interstellar Faraday rotation reveals strong disturbances imposed on the regular field by magnetic bubbles associated with supernova remnants and stellar winds^{24,25}. Another factor that introduces irregularity to galactic magnetic fields is the fluctuating magnetic field associated with interstellar turbulence. The r.m.s. strength of magnetic field fluctuations in the Galaxy is 1–2 times greater than the regular field strength and their correlation scale is approximately the same as the energy-range scale of interstellar turbulence, $l \approx 50\text{--}150$ pc. Chaotic fields of comparable strength are also observed in other spiral galaxies^{20,21}.

In the past ten years, systematic observations using the 100-m radiotelescope in Effelsberg, Germany have constituted a major step forward in the study of galactic magnetism, making possible the investigation of global configurations of magnetic fields in nearby spiral galaxies^{20,21,26}. Previous observations were restricted to the solar neighbourhood in the Milky Way, leaving much ambiguity about global structure of the magnetic field. Observations of nearby galaxies are restricted by angular resolution (somewhat relieved by recent interferometric observations with the VLA²⁶) and they still leave unanswered many questions about the z -symmetry and chaotic components of the fields.

Yakov Borisovich Zel'dovich (8.3.1914–2.12.1987) was one of the guiding lights of modern physics. He made essential contributions to the theory of nuclear fission and, in later years, to astrophysics and cosmology, including work on the cosmological constant, black holes and neutron stars, galaxy formation and cosmic microwave background radiation.



Theoretical prediction²⁷ of a magnetic ring in the galaxy M31 (the Andromeda Nebula) was confirmed almost immediately by observations at Effelsberg^{28,29}. Observations³⁰ of another nearby spiral galaxy, M51, were interpreted³¹ as indicating a basically non-axisymmetric global magnetic configuration, with magnetic lines seemingly entering the galactic disk from one side and leaving it from another. Such a configuration is called bisymmetric³¹. Bisymmetric magnetic structures were a direct challenge to the theory of galactic magnetism^{32,33}.

Global magnetic structures have now been studied for a dozen nearby spiral galaxies. These structures are crudely divided into two basic configurations, axisymmetric and bisymmetric^{20,21}. The majority of galaxies appear to possess bisymmetric fields, whereas only a few (M31, IC342 and, probably, the Milky Way) exhibit axisymmetric structures. But the dynamo theory predicts that combinations of such basic modes must be common, including those modes with still more complicated symmetries relative to the azimuthal angle. The estimated strengths of regular fields in different spiral galaxies range from 1 to 5 μG and the ratio of fluctuations to the regular field ranges from 0.4 to 2.5.

Contrary to earlier views, recent theories, confirmed by observations²⁶, stress that the field lines of large-scale magnetic fields in all structures are always spirals rather than closed circles^{34,35}.

There are many directions in which these observations can be further developed. Observations with higher angular resolution should provide new insights. There is much scope for improvement of data processing methods. Careful determinations of intrinsic Faraday rotation measure distributions are needed. Apparent dominance of bisymmetric fields among observed structures should be reconsidered for more representative samples and with better data processing. What is important, however, is the fact that the existence of non-axisymmetric regular fields seems to be firmly established. The discovery of various global magnetic structures in spiral galaxies has provoked interest in the origin of galactic magnetism.

The origin of galactic magnetic fields

Hoyle³⁶ attracted attention to the origin of the galactic magnetic field by pointing out that an enormous electromotive force ($\sim 10^{12}$ V) must act in the ionized galactic disk for the entire galactic lifetime to build up and maintain the observed large-scale field. This forced him to suggest that the field is a relic from pre-galactic stages of evolution of the Universe, requiring the presence of cosmological magnetic fields whose scales exceed the galactic one. Some ideas proposed to explain how these fields could be generated in the Universe are discussed in ref. 18. All proposed mechanisms give only very weak present-day intergalactic magnetic fields of $H_{\text{IG}} = 10^{-9}\text{--}10^{-21}$ G at galactic scales. An upper limit on cosmological magnetic fields is provided by analysis of the observed Faraday rotation of polarized radio emission of remote radio sources (quasars and radio

galaxies): $H_{IG} \approx 5 \times 10^{-10}$ G for $n_e = 10^{-5} \text{ cm}^{-3}$ in intergalactic space³⁷ (estimates of the latter parameter remain uncertain³⁸). Primordial fields of below 10^{-8} – 10^{-9} G are insufficient to explain the observed galactic magnetic fields³⁹. But even if the cosmological magnetic field were of the 10^{-8} G strength required by the relic-origin hypothesis, it could not account for the observed large-scale magnetic fields in spiral galaxies because the conductive interstellar gas is in a state of intense turbulence that inevitably tangles away any large-scale magnetic field unless this is constantly replenished by some other mechanism⁴⁰. The energy-range scale and velocity of the interstellar turbulence are $l \approx 100$ pc and $v \approx 10 \text{ km s}^{-1}$ respectively, which gives an estimate of $\beta \approx lv/3 \approx 10^{26} \text{ cm}^2 \text{ s}^{-1}$ for the turbulent magnetic diffusivity. Therefore, in galactic disks of half-thickness $h \approx 400$ pc the relic large-scale magnetic field would be destroyed in $h^2/\beta \approx 5 \times 10^8$ years, which spans only five per cent of the Galactic lifetime.

Thus, a relic origin of magnetic fields of spiral galaxies can be dismissed on both observational and theoretical grounds. Nevertheless, recent detection of non-axisymmetric global magnetic structures in nearby galaxies has resurrected suggestions of their relic origin on the grounds of their outward resemblance to the magnetic structures produced by differential rotation of galaxies from a uniform external magnetic field lying in the galactic plane. But the differential rotation winds up such magnetic fields very quickly (one rotation typically takes only 3×10^8 years) which enhances dissipation of the trapped field through decrease in its radial scale. Observations show that magnetic lines of the non-axisymmetric magnetic fields have moderate pitch angles to the circumferential direction²⁶, of order 10° – 20° . This precludes the possibility that large-scale magnetic fields are passive remnants of an intergalactic field and consequently they must originate within galaxies.

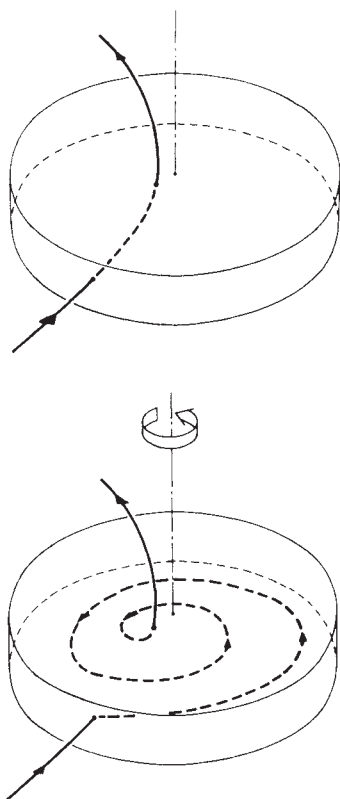


Fig. 2 The non-uniform rotation of spiral galaxies turns a purely meridional field into one which is a combination of meridional and azimuthal components. The angular velocity of the gas is generally higher at smaller galactic radii, and a trapped meridional field (top) is stretched into a trailing spiral (bottom).

It is now commonly accepted that galactic magnetic fields are generated by the action of a turbulent hydromagnetic dynamo, that is, by motions of the ionized interstellar gas. The idea that helical turbulent motions are able to generate large-scale magnetic fields was proposed by Parker⁴¹ and became commonly accepted after the decisive work of Steenbeck, Krause and Rädler⁴². The dynamo theory was first applied to spiral galaxies by Parker⁴³ and Vainshtein and Ruzmaikin⁴⁴. We consider now how the galactic dynamo works.

The galactic dynamo

In the evolution of magnetic fields embedded in interstellar gas, the most important characteristic of the medium is the magnetic Reynolds number Re_m , which measures the role of the ionized gas motions relative to the role of magnetic diffusion:

$$Re_m = \frac{VL}{\nu_m}$$

where V is the characteristic gas velocity, L is the velocity-field scale and ν_m is the magnetic diffusivity, $\nu_m = c^2/4\pi\sigma$, with σ the electric conductivity. Although conductivity of interstellar gas is low relative to that of metallic conductors, the galactic scales involved are so large that the magnetic Reynolds number is very high in the interstellar environment. For the values $V = 10 \text{ km s}^{-1}$, $L = 100$ pc, which characterize the energy-range scale of interstellar turbulence, and $\nu_m \approx 2 \times 10^{21} \text{ cm}^2 \text{ s}^{-1}$ (this being the ambipolar diffusion in a partially ionized interstellar hydrogen plasma at $T = 10^4$ K, $n = 1 \text{ cm}^{-3}$, for the total field strength $H \approx 5 \mu\text{G}$; note that $\nu_m \propto H^2$), we obtain $Re_m \approx 5 \times 10^4$. This very high value indicates that interstellar magnetic fields can be conveniently considered as frozen into the interstellar gas unless the scales involved are below 10^{-1} – 10^{-2} pc (refs 45, 46).

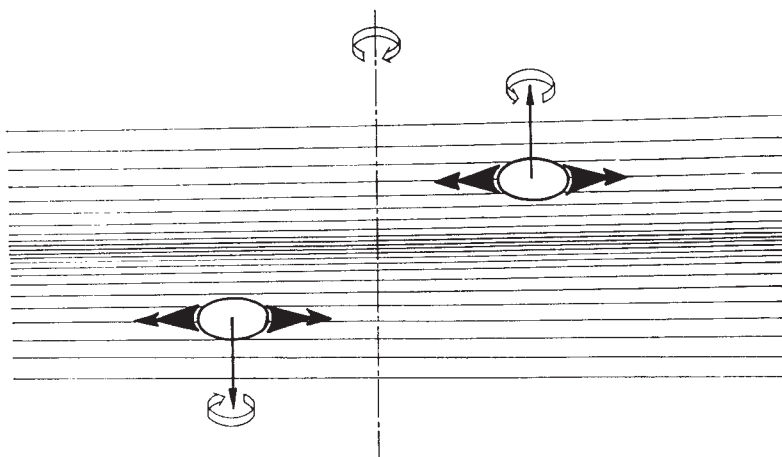
In the dynamo theory, the term 'generation' is always understood as an exponential time-growth of magnetic flux. At the initial stage of exponential growth the velocity field may be considered as fixed and the dynamo is called kinematic. Any magnetic field can be decomposed into azimuthal (B_ϕ) and meridional (B_r and B_z) components. It is clear that the field can grow exponentially only when both components are mutually coupled, that is, when there is a way of transforming the meridional component into the azimuthal one and vice versa. This situation is typical of any instability. Because the induction equation is linear in $\mathbf{B}$, then if one component were fixed the other could grow at best linearly with time.

The first stage of the galactic dynamo process, the effect of differential rotation, is illustrated in Fig. 2. The feedback loop of the galactic dynamo is associated with a more subtle feature of the motions of the interstellar gas, the mean helicity of the turbulence (Fig. 3). Deviations of the averaged characteristics from reflection symmetry are exceptional in laboratory turbulent flows but they are a quite general feature of cosmic environments where turbulent volumes rotate and are density-stratified. The average deviation from reflection symmetry in the presence of an azimuthal field results in emergence of a mean meridional magnetic field (Fig. 4).

We have illustrated the galactic dynamo mechanism for magnetic fields frozen into the interstellar medium, but finite magnetic diffusivity is necessary for galactic mean-field dynamos¹⁸. We note that the possibility of a dynamo arising in ideally conducting medium is still under debate: it is traditionally believed that a dynamo cannot exist in finite bodies without resistivity, but there exists an explicit example of a dynamo for $\nu_m = 0$, this being the case of renovating flow in infinite space⁴⁷.

Generation of the meridional field implies a growing azimuthal electric current. In other words, the action of helical turbulence is equivalent to the generation of an azimuthal electromotive force αB_ϕ where α is a function of position and is proportional to the mean helicity of the turbulence. Therefore, the induction equation, when averaged over turbulent pulsa-

Fig. 3 A schematic vertical section through a galactic disk illustrating breaking of the reflection invariance of the turbulent flow, characterized mathematically by the scalar product $\mathbf{v} \cdot \nabla \times \mathbf{v}$, called the helicity. A turbulent cell rising up through the disk expands because the density scale height of the interstellar gas only slightly exceeds the correlation scale of the turbulence. The Coriolis force from galactic rotation delivers an additional rotation to the expanding cells. Descending cells contract and acquire an oppositely directed Coriolis rotation, but all cells above (or below) the galactic plane have helicity of the same sign. The mean helicity $\langle \mathbf{v} \cdot \nabla \times \mathbf{v} \rangle$, where $\langle \dots \rangle$ denotes an ensemble average, is thus an odd function of the vertical coordinate z , and breaks the mirror symmetry of the galactic plane.



tions, acquires an additional term associated with this e.m.f.:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\alpha \mathbf{B}) + \nabla \times (\mathbf{V} \times \mathbf{B}) + \beta \Delta \mathbf{B} \quad (1)$$

where $\mathbf{B} = \langle \mathbf{H} \rangle$, $\mathbf{V}$ is the regular velocity and $\alpha = -\tau(\mathbf{v} \cdot \nabla \times \mathbf{v})/3$ with τ the correlation time and $\mathbf{v}$ the turbulent velocity. Mathematically rigorous averaging procedures for the induction equation are described for various conditions in refs 18 and 47–49.

Equation (1) describes the generation of magnetic fields not only in galaxies but also in stars and planets. Differences in geometries of these objects result in drastic differences in properties of the generated fields: in disk galaxies they are quadrupolar and non-oscillating in a rotating frame, in a thin convective shell of the Sun they are dipolar and oscillatory whereas a thick spherical shell of the Earth's outer core generates a dipolar quasi-stationary field^{17,48}.

Self-induced fields in galactic disks

In a spiral galaxy the interstellar gas is concentrated predominantly within a very thin disk. For example, in the solar vicinity of the Milky Way the ratio of the half-thickness of the disk of ionized gas to the disk radius is

$$\lambda = \frac{h}{r} \approx \frac{400 \text{ pc}}{10 \text{ kpc}} = 0.04 \quad (2)$$

The disk thickness grows somewhat towards the edge. Of course, in central parts ($r \approx h$) the disk cannot be considered thin and the solution of the dynamo problem is then a separate problem⁵⁰.

As explained above, the large-scale magnetic field in the disk of a spiral galaxy is produced by competition between regenerative action of the mean helicity and differential rotation, and attenuation by turbulent diffusion. The diffusion time across a thin disk is obviously much shorter than that along the radius or azimuth. This allows us first to solve the local generation problem and to deduce the field distribution across the disk, and then to use this solution as a basis for evaluation of the radial and azimuthal distributions of the field.

The strengths of the two generation sources—the mean helicity and differential rotation—that compete with turbulent diffusion are measured by the respective dimensionless numbers,

$$R_\alpha = \frac{\alpha_0 h_0}{\beta} \quad \text{and} \quad R_\omega = \frac{|r d\Omega/dr|_0 h_0^2}{\beta}$$

where subscript zero denotes characteristic values of the corresponding quantities and Ω is the angular velocity of rotation.

In the kinematic formulation of the dynamo problem, the generation sources are considered as fixed and therefore uninfluenced by back-action of the generated field. This formulation

is adequate for determination of the field origin and is applicable over most of the galactic lifetime. When generation sources have fixed strengths the dynamo equation (1) has solutions that evolve exponentially with time and for small R_α and R_ω the solutions decay. For positive growth rate of the field (that is, self-excitation) the following necessary condition must be fulfilled:

$$R_\alpha R_\omega \geq D_c$$

where D_c , the critical dynamo number, is of order 10 (see below).

The solution of the dynamo equation (1) in an axisymmetric disk with $\mathbf{V} = \Omega \times \mathbf{r}$ has the form

$$\mathbf{B} = \mathcal{B}(r, z) \exp(\Gamma t + im\phi)$$

where r is the distance to the rotation axis and m is the azimuthal wave number. It is convenient to introduce dimensionless variables by measuring the vertical coordinate in units $h_0 = 400 \text{ pc}$, the radius in units $r_0 = 10 \text{ kpc}$, and the time in units of the diffusion time $h_0^2/\beta \approx 5 \times 10^8 \text{ years}$. The mean helicity and the angular velocity are conveniently normalized by $\alpha_0 = 1 \text{ km s}^{-1}$ and $\Omega_0 = 25 \text{ km s}^{-1} \text{ kpc}^{-1}$. The growth rate Γ is measured in units of β/h_0^2 .

Equation (1) is essentially a system of three linear second-order differential equations. For uniform coefficients (the sources α and $r d\Omega/dr$) this system can be solved analytically. But in galaxies these coefficients depend on position, and numerical solutions are then appropriate. We can make a crude estimate of the computer memory required for a straightforward solution, which indicates that vectors with 10^9 components and matrices with 10^{18} elements must be stored. This impractical result is associated with the presence of the small parameter λ , which makes the intrinsic geometry very inconvenient for straightforward solution; for cubic geometry, some 3×10^3 mesh points would suffice to reach the same accuracy. Some progress can be made with numerical methods through the presence of certain symmetries, judicious choice of the mesh, restriction to simplified models of galaxies, and other factors^{51,52}.

The smallness of λ can, however, be used to advantage. When $\lambda \ll 1$, one can employ a simple asymptotic method (in combination with some numerical calculations) to find the growth rate eigenvalue Γ and the eigenfunctions $\mathcal{B}(r, z)$ for suitably chosen boundary conditions and for distributions of the generation sources inferred from observations. The boundary conditions routinely used in galactic dynamo models assert that the galactic disk is surrounded by vacuum^{17,18,40,48}. Formulation of more adequate boundary conditions would be desirable because galactic disks are in fact surrounded by conducting coronae. Nevertheless, the vacuum boundary conditions are partially justified by the observation that, because of higher turbulent scales and velocities in galactic coronae relative to the disk, the turbulent magnetic diffusivity is expected to be some 30 times higher in the corona.

In this asymptotic solution the radial scale of the field is assumed to be large, of order $\lambda^{-1/2}$, so that the radial derivatives of the dimensionless field components are of order $\lambda^{1/2}$ whereas the vertical derivatives are of order λ (refs 34 and 53). With such ordering, and retaining only the terms up to order λ^2 , equation (1) reduces to

$$\left\{ \Gamma + imR_\omega \Omega(r) - \frac{\partial^2}{\partial z^2} - \lambda^2 \frac{\partial}{\partial r} \left[\frac{1}{r} \frac{\partial}{\partial r} (r \times \cdot) \right] \right\} \begin{bmatrix} \mathcal{B}_r \\ \mathcal{B}_\phi \end{bmatrix} = \begin{bmatrix} 0 & -R_\alpha \frac{\partial}{\partial z} (\alpha \times \cdot) \\ R_\omega r \frac{d\Omega}{dr} & 0 \end{bmatrix} \begin{bmatrix} \mathcal{B}_r \\ \mathcal{B}_\phi \end{bmatrix} \quad (3)$$

At this order, the equation for the vertical field component, B_z , factorizes out and may be solved separately. In equation (3) we have neglected a term in R_α in the evaluation of $\mathcal{B}_\phi$, as it is generally somewhat smaller than the R_ω term (typically, $R_\omega \approx 10$ and $R_\alpha \approx 1$ in galactic disks) and is not qualitatively significant in a thin disk⁵⁴. But in some galaxies (for example, M33 and NGC6946) the large-scale field is generated at rather small distances from the centres, where the disk cannot be considered thin; in these cases inclusion of the R_α term is essential⁵⁰.

To lowest order in λ , equation (3) has a solution of the form $\mathcal{B}(r, z) = Q(r)\mathbf{b}(z)$, where $Q(r)$ is a radial function and $\mathbf{b}(z)$ is the matrix of local radial (b_r) and azimuthal (b_ϕ) components of the field. These local components satisfy one-dimensional 'local dynamo equations' with eigenvalues $\gamma(r)$ that represent local growth rates of the field. Vertical and radial distributions of the field will be discussed separately.

Vertical distribution of the field. For conditions typical of spiral galaxies the local eigenvalue $\gamma(r)$ is real. The local dynamo equation has growing solutions $\gamma(r) > 0$ only for sufficiently strong generation sources. Specifically, the product $D_{\text{eff}}(r) \equiv R_\alpha R_\omega h^3 |\alpha r d\Omega/dr|$, the local dynamo number, must exceed a certain threshold value that depends on the distribution of the mean helicity (or α) across the disk. For instance, for $\alpha = \sin(\pi z)$ this threshold is $D_{\text{cr}} \approx 8$ while for $\alpha = z$, $D_{\text{cr}} \approx 11$.

In the principal magnetic mode b_ϕ and b_r are even functions of z . In addition, the azimuthal field turns out to be considerably larger than the other components, $b_r/b_\phi = O(R_\alpha/R_\omega)^{1/2}$ (refs 17, 18) (here, local values of R_α and R_ω should be used). The field runs almost parallel to the galactic plane and has a small pitch angle with respect to the circumferential direction. All these results agree favourably with observations of magnetic fields in the solar vicinity.

Because the dynamo equations have only decaying solutions when either of the field components identically vanishes, growing solutions can never have magnetic lines that form closed circles in the galactic plane. Magnetic lines of the dynamo-generated fields are always spirals. The radial and azimuthal field components have different signs, which means that the spirals are trailing.

The azimuthal wave number m does not enter the local generation problem. This implies that modes with different m are distributed over z nearly identically. Furthermore, pitch angles of magnetic lines are approximately the same for all m ; this is also confirmed by observations²⁶.

Radial distribution of the field. It is easily shown that the growing magnetic fields are concentrated within those radial ranges where $\gamma(r) > 0$. Moreover, the global growth rate $\text{Re}(\Gamma)$ can be positive only when the potential well is sufficiently wide and deep, that is, when $D_{\text{eff}}(r) > D_{\text{cr}}$ over a sufficiently wide radial range. The value of $\text{Re}(\Gamma)$ can be considerably lower than the maximal value of γ ; for instance, $\max(\gamma) \approx (5 \times 10^8 \text{ yr})^{-1}$ (ref. 17) whereas $\text{Re}(\Gamma) \approx (1.5 \times 10^9 \text{ yr})^{-1}$ (ref. 54) in the Milky Way. Within an order of magnitude, $\mathcal{B}_z \approx \lambda^{1/2} \mathcal{B}_r \approx (\lambda R_\alpha/R_\omega)^{1/2} \mathcal{B}_\phi$.

Analysis of the local generation problem shows that under

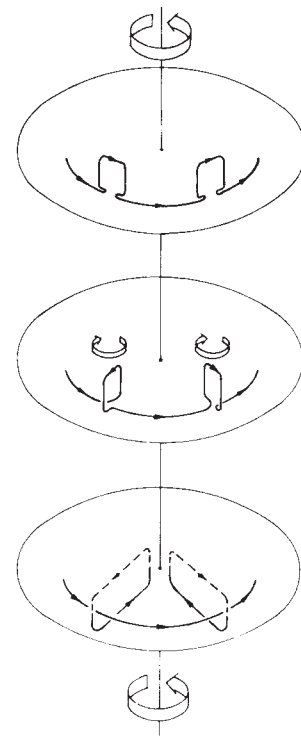


Fig. 4 Helical turbulence in a conducting fluid creates a meridional magnetic field from an azimuthal one. Turbulent cells locally distort the azimuthal magnetic line (top); because the turbulence is helical, local magnetic loops rotate about the vertical axis (middle). Oppositely directed fields at the base of each loop are quickly removed by electric resistivity, and small loops detach from the parent azimuthal magnetic line to merge, again due to electric resistivity which is important at small scales. In helical turbulence the total numbers of oppositely twisted loops are not identical and after their merging the small-scale fields are not cancelled completely but produce a large-scale meridional magnetic line (bottom). The upper part of the resulting line (dashed) leaves the disk. The meridional field thus growing within the disk has the direction required for further amplification of the azimuthal component by differential rotation (compare Fig. 2). Helicity can also produce an azimuthal field from a meridional one but in spiral galaxies this process is generally weak compared with a similar action of the differential rotation.

galactic conditions field generation is possible only when $d\Omega/dr < 0$; when the angular velocity grows with r the critical dynamo number is very high, of the order of a few hundreds. This fact explains why the magnetic field in the galaxy M31 is concentrated within a ring around $r = 10$ kpc and in the central region: these domains are separated by a gap in the region $2 \leq r \leq 6$ kpc, for which $d\Omega/dr > 0$ and where consequently a field cannot be maintained. Magnetic field generation proceeds independently in the outer and inner regions. This gap in the magnetic field distribution of M31 was predicted theoretically²⁷ and observed with the Effelsberg radiotelescope²⁹.

Although in the Milky Way $d\Omega/dr < 0$ for all r , in the region $3 \leq r \leq 6$ kpc rotation is relatively close to rigid-body rotation (see, for example, ref. 55). Therefore, $\gamma(r)$ may become a double-well potential and the generated field will be then concentrated in an outer ring, coincidentally near $r = 10$ kpc, and in the central part of the disk. But the value of D_{eff} depends not only on $d\Omega/dr$ but also on the ionized disk half-thickness h , the value of which is uncertain. For sufficiently large disk thickness, $D_{\text{eff}} > D_{\text{cr}}$ for all r , in which case there are no gaps in the magnetized region³⁴. Thus, from the field distribution one can infer an estimate of the galactic disk thickness.

Dynamo modes with different azimuthal wave numbers behave very differently, because differential rotation creates distinct generation conditions for different m -modes. Generation conditions are most favourable for the axisymmetric mode because its radial and azimuthal scales are the largest. Radial and azimuthal diffusion is more important for non-axisymmetric fields, which have smaller growth rates³². In the absence of a dynamo, non-axisymmetric magnetic fields would be expelled from regions with strong differential rotation; similarly, dynamo modes with $m \neq 0$ concentrate at larger radii, where the local value of R_ω is smaller, than axisymmetric ones. An important prediction of the galactic dynamo model is the presence of combinations of axisymmetric and non-axisymmetric fields in spiral galaxies³⁴, indications of which have been detected in the Andromeda Nebula²⁹.

The current theory of galactic dynamos is in general agreement with observations; theoretical models exist for nearly all galaxies whose large-scale magnetic fields have been observed in any detail^{34,50,53,54}. For example, Fig. 5 shows magnetic lines of the growing dynamo modes $m=0$ in M31 and $m=1$ in M51 as determined⁵⁴ from the theory outlined earlier. Observations, however, seem to indicate a much wider abundance of non-axisymmetric magnetic structures among spiral galaxies than is expected from theory: only two out of a dozen galaxies investigated definitely host axisymmetric fields. This observation constitutes one of the intriguing current problems in galactic magnetism. The theory described above considers generation of a field in an axially symmetric disk, and it is therefore no surprise that axisymmetric magnetic modes always have the greatest growth rates in this model. Why, then, do non-axisymmetric modes dominate amongst observed magnetic structures? Deviations of galactic disks from exact axial symmetry, associated with the presence of spiral arms, would facilitate generation of bisymmetric fields. But these perturbations correspond to $m=2$ (two-armed spiral patterns dominate) and therefore they cannot affect generation conditions for the $m=0$ and $m=1$ modes, at least at the level of the weak-perturbation analysis. The observed dominance of the $m=1$ modes in spiral galaxies may be the result of corresponding $m=1$ non-axisymmetric perturbations⁵⁴ associated with tidal interactions with companion galaxies and with warpings of galactic disks. The connection between galactic satellites and bisymmetric fields is discussed on observational grounds in ref. 56. Another possible explanation is provided by anisotropy of interstellar turbulence⁵⁴.

Amongst other problems associated with the galactic dynamo theory we mention an urgent need for more precise observational determination of the thickness, as a function of radius, of the ionized gas layers in spiral galaxies and for the measurement of the mean helicity of interstellar turbulence and its distribution across the disk. Helical turbulent flows are also crucial for generation of magnetic fields in planets and stars, so their detection in galaxies plays a fundamental role. Interstellar gas, being a unique dynamo region transparent to optical and radio waves, offers an exceptional opportunity to verify fundamental tenets of the dynamo theory.

Seed fields

Like other dynamo mechanisms, the mean-field turbulent dynamo requires some initial large-scale magnetic field. If we reject the hypothesis of a relic intergalactic magnetic field we are forced to propose a galactic-scale seed field and must then address the problem of its origin. Battery effects are effective only at small scales and can produce only exceedingly weak seed fields at the relevant larger scales¹⁸.

An unexpected solution to the problem may be found in the magnetic fields ejected from stars by supernova explosions and by stellar winds. Such ejections yield a chaotic field whose value is zero when averaged over an infinitely large and homogeneous volume. But galactic disks have finite volumes that can contain, within the region of field generation, perhaps 300 or so of the

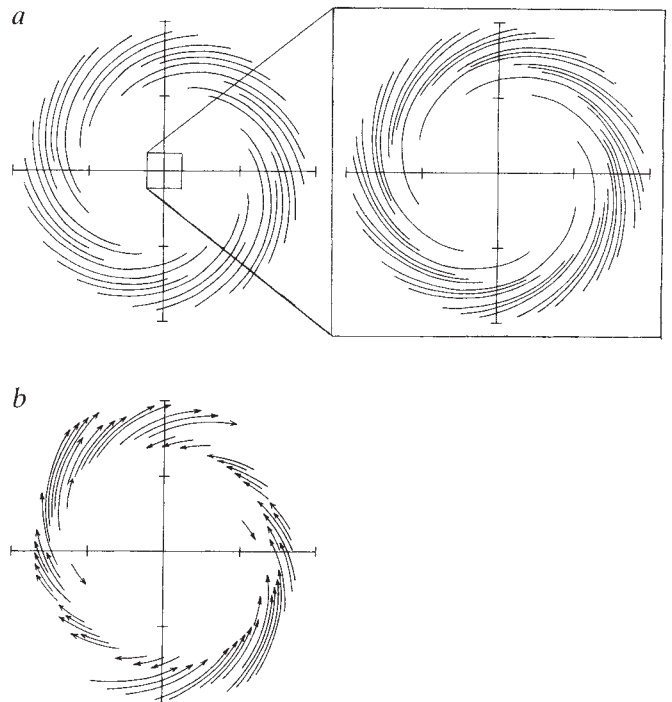


Fig. 5 Magnetic flux lines in *a*, the Andromeda galaxy, M31, and *b*, M51 are shown here projected onto their respective galactic planes. The magnetic field in M31 is generated both in the outer ring and near the centre. The field structures are dominated by the $m=0$ mode in M31 and possibly the $m=1$ mode in M51. Because the kinematic dynamo equations are invariant under $\mathbf{B} \rightarrow -\mathbf{B}$, the directions of the field lines in *a* are arbitrary, provided only that the direction at all azimuthal angles is uniform. Likewise, the arrows in *b* may all be reversed.

100-pc magnetic loops that characterize such ejections. The ejected field averaged over such volumes may be significantly different from zero. Thus the long-range order originates from a fluctuation and is stabilized by the dynamo. The importance of fields ejected by stars in the global galactic context was noted in refs 57 and 58, but an adequate formulation of the problem was developed only recently³² (see also the discussion by Rees⁵⁹).

To examine the evolution of a large-scale magnetic field that originates in this way, one can expand the ejected seed field $\mathbf{B}_0(\mathbf{r})$ over eigenfunctions $\mathbf{B}(\mathbf{r})$ of the kinematic dynamo problem. Magnetic lines of the field $\mathbf{B}_0$ are small loops closed in correlation cells. Thus the field $\mathbf{B}_0$ oscillates rapidly with a characteristic scale $l \approx 100$ pc, much smaller than the scales of the dynamo modes. Because the loops are placed randomly, the oscillations cannot be described by a periodic function but are instead characterized as an ensemble of derivatives of δ -functions scattered randomly in space (Fig. 6). This model preserves the closed-loop solenoidality of the small-scale magnetic field. By approximating the lowest axisymmetric eigenfunction as a radially uniform field directed along the azimuth and with radial width Δr , one can integrate the scalar product of $\mathbf{B}(\mathbf{r})$ with $\mathbf{B}_0(\mathbf{r})$ over the volume of the galactic disk to give an estimate of the initial (seed) amplitude of the large scale field as:

$$B_\phi \Big|_{t=t_0} \approx \frac{bl}{N^{1/2}\Delta r}$$

where b is the r.m.s. strength of the random field ejected by stars and N is the number of correlation cells within the localization region of the dynamo mode. If we use the typical values $b \approx 1 \mu\text{G}$, $N \approx 300$ and $\Delta r \approx 3$ kpc, we obtain a seed amplitude of $2 \times 10^{-3} \mu\text{G}$. A growth rate of $(1.5 \times 10^9 \text{ yr})^{-1}$ is required to

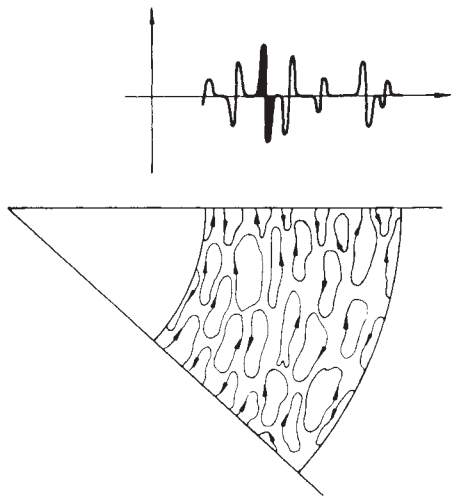


Fig. 6 The chaotic magnetic field ejected by stars can be approximated, along any radial slice, by a collection of derivatives of δ -functions. This model preserves solenoidality of the chaotic magnetic field and allows an estimation of the seed field for the galactic dynamo.

amplify this field up to the observed strength of $2 \mu\text{G}$, which is in excellent agreement with estimates of the dynamo theory. Magnetic fields of first-generation stars formed in non-magnetized interstellar gas can be produced by battery mechanisms⁶⁰.

Nonlinearity

The exponential growth of the magnetic field described by the kinematic dynamo saturates when the nonlinear influence of magnetic fields on the generating motions becomes important. Because differential rotation (R_ω) is a stronger generator (by orders of magnitude) than the mean helicity of turbulence (R_α), it is clear that generation is suppressed because of modification of the mean helicity. The helicity is produced by the Coriolis force, so it is expected that the field approaches a steady state when magnetic stresses balance the Coriolis force:

$$\frac{B_r B_\phi}{4\pi h} \approx \rho v \Omega,$$

where h is the minimal field scale and ρ is the gas density.

Because $B_r/B_\phi \approx O(R_\alpha/R_\omega)^{1/2}$ (refs 17, 18), one obtains

$$B_\phi \approx \left[\frac{R_\alpha}{R_\omega} \right]^{-1/4} (4\pi\rho v h \Omega)^{1/2}$$

For $h \approx 400 \text{ pc}$, $\rho \approx 10^{-24} \text{ g cm}^{-3}$, $v \approx 10 \text{ km s}^{-1}$, $\Omega \approx 10^{-15} \text{ s}^{-1}$, $R_\alpha \approx 1$ and $R_\omega \approx 10$, this estimate yields $B_\phi \approx 7 \mu\text{G}$, which is of the same order as the observed strength of the large-scale field. Accompanying estimates are $B_r \approx 0.3 B_\phi$ and $B_z \approx \lambda^{1/2} B_r \approx 0.05 B_\phi$. We emphasize that these estimates are valid only as an average over the disk. In some regions the local ratios of the field components can be very different. For instance, in the centre of a galaxy, where λ effectively approaches unity, the vertical field becomes comparable to other field components. Magnetic fields at distances of several parsecs from galactic centres can be generated by a turbulent dynamo independently of the global field^{18,23}.

In the galactic dynamo, nonlinearity serves mainly as a saturation factor. Such complex, essentially nonlinear, global magnetic phenomena as irregular inversions of the Earth's magnetic field or global activity minima in the Sun (for example, the Maunder minimum) simply never have time to develop in spiral galaxies, because the typical growth time of the large-scale magnetic field (1.5×10^9 years for the Milky Way) is only slightly shorter than the galactic lifetime (10^{10} years). Large-scale galactic magnetic fields have sufficient time to grow from an initial $10^{-3} \mu\text{G}$ ejected by stars only up to the present-day strength. In this respect, spiral galaxies are rather young. Nonlinear effects certainly play a role in galactic dynamos. This can be seen, for example, in the fact that the radial widths of magnetic field distributions observed in spiral galaxies are significantly larger than those of the kinematic dynamo eigenfunctions, which is explicable in terms of nonlinear dynamo theory³². More varied nonlinear effects can be expected at smaller, turbulent scales.

Galaxies may differ considerably in their magnetic ages. Comparable large-scale field strengths in different galaxies indicate that galactic dynamos are at or near saturation. In addition, the characteristic time is much shorter, and nonlinear effects are therefore much stronger, in the centre of the Milky Way and, presumably, other galaxies. But in the solar vicinity of the Milky Way, where the dynamo is comparatively weak, we witness a unique epoch, of duration $\sim 1.5 \times 10^9$ years, when its large-scale dynamo is just entering the nonlinear stage.

- Prigogine, I. & Stengers, I. *Order out of Chaos. Man's New Dialogue With Nature* (Heinemann, London, 1984).
- Fermi, E. *Phys. Rev.* **171**, 1169–1174 (1949).
- Alfvén, H. *Z. Phys.* **105**, 579–588; *Z. Phys.* **107**, 319–333 (1937).
- Richtmeier, R. D. & Teller, E. *Phys. Rev.* **75**, 1729–1731 (1949).
- Hall, J. S. & Mikesell, A. M. *Astrophys. J.* **54**, 187–188 (1949).
- Hiltner, W. A. *Nature* **163**, 283 (1949).
- Alfvén, H. & Herlofson, N. *Phys. Rev.* **78**, 616 (1950).
- Kiepenheuer, K. O. *Phys. Rev.* **79**, 738–739 (1950).
- Ginzburg, V. L. *Usp. fiz. Nauk (USSR)* **51**, 343–392 (1953).
- Shklovsky, I. S. *The Cosmic Radioemission* (Gostekhizdat, Moscow, 1956).
- Wielebinski, R., Shakeshaft, J. R. & Pauliny-Toth, I. I. K. *Observatory* **82**, 158–164 (1962).
- Westerhout, G., Seeger, Ch. L., Brouw, W. N. & Tinbergen, J. *Bull. astr. Inst. Netherlands* **16**, 187–212 (1962).
- Razin, V. A. *Sov. Astr.* **2**, 216–225 (1958).
- Heiles, C. A. *Rev. Astr. Astrophys.* **14**, 1–22 (1976).
- Spoelstra, T. A. Th. *Sov. Phys. Usp.* **20**, 336–342 (1977).
- Verschuur, G. L. *Fundam. Cosmic Phys.* **5**, 113–191 (1979).
- Zel'dovich, Ya. B., Ruzmaikin, A. A. & Sokoloff, D. D. *Magnetic Fields in Astrophysics* (Gordon & Breach, New York, 1983).
- Ruzmaikin, A. A., Shukurov, A. M. & Sokoloff, D. D. *Magnetic Fields of Galaxies* (Kluwer Academic Publ., Dordrecht, 1988).
- Rand, A. & Kulkarni, S. R. *Astrophys. J.* (submitted).
- Beck, R. *IEEE Trans. Plasma Sci.* **PS-14**, 740–747 (1986).
- Sofue, Y., Fujimoto, M. & Wielebinski, R. A. *Rev. Astr. Astrophys.* **24**, 459–498 (1986).
- Sofue, Y. *Publs astr. Soc. Japan* **39**, 547–558 (1987).
- Yusef-Zadeh, F. & Morris, M. *Astrophys. J.* **320**, 545–561 (1987).
- Spoelstra, T. A. Th. *Astr. Astrophys.* **21**, 61–84 (1972).
- Vallée, J. P. *J. R. astr. Soc. Can.* **78**, 57–74 (1984).
- Interstellar Magnetic Fields* (eds Beck, R. & Gräve, R.) (Springer, Berlin, 1987).
- Ruzmaikin, A. A. & Shukurov, A. M. *Sov. Astr.* **25**, 553–558 (1981).
- Sofue, Y. & Takano, T. *Publs astr. Soc. Japan* **33**, 47–55 (1981).
- Beck, R. *Astr. Astrophys.* **106**, 121–132 (1982).
- Segalovitz, A., Shane, W. W. & de Bruyn, A. G. *Nature* **264**, 222–226 (1976).
- Tosa, M. & Fujimoto, M. *Publs astr. Soc. Japan* **30**, 315–325 (1978).
- Ruzmaikin, A. A., Sokoloff, D. D. & Shukurov, A. M. in *Plasma Astrophysics, Proc. Varenna-Abastumani Workshop* 539–544 (ESA Publ., SP-251, 1986).
- Sawa, T. & Fujimoto, M. *Publs astr. Soc. Japan* **38**, 133–150 (1986).
- Baryshnikova, Yu., Ruzmaikin, A., Sokoloff, D. & Shukurov, A. *Astr. Astrophys.* **177**, 27–41 (1987).
- Ruzmaikin, A. in *Interstellar Magnetic Fields* (eds Beck, R. & Gräve, R.) 16–22 (Springer, Berlin, 1987).
- Hoyle, F. in *La structure et l'évolution de l'univers: XI conseil de physique Solvay* (ed. Stoops, R.) 53–73 (1958).
- Ruzmaikin, A. A. & Sokoloff, D. D. *Astr. Astrophys.* **58**, 247–253 (1977).
- Zel'dovich, Ya. B. & Novikov, I. D. *The Structure and Evolution of the Universe* (Univ. Press, Chicago, 1982).
- Piddington, J. H. *Astrophys. & Space Sci.* **80**, 457–471 (1981).
- Parker, E. N. *Cosmical Magnetic Fields*, Sect. 22.2.1 (Clarendon Press, Oxford, 1979).
- Parker, E. N. *Astrophys. J.* **122**, 293–314 (1955).
- Steenbeck, M., Krause, F. & Rädler, K.-H. *Z. Naturf.* **21a**, 369–376 (1966).
- Parker, E. N. *Astrophys. J.* **163**, 255–278 (1971).
- Vainshtein, S. I. & Ruzmaikin, A. A. *Astr. J. (USSR)* **48**, 902–909 (1971).
- Melvor, I. *Mon. Not. R. astr. Soc.* **178**, 85–100 (1977).
- Ruzmaikin, A. A. & Shukurov, A. M. *Astrophys. & Space Sci.* **82**, 397–407 (1982).
- Molchanov, S. A., Ruzmaikin, A. A. & Sokoloff, D. D. *Sov. Phys. Usp.* **28**, 307–327 (1985).
- Moffatt, H. K. *Magnetic Field Generation in Electrically Conducting Fluids* (Cambridge Univ. Press, Cambridge, 1978).
- Krause, F. & Rädler, K.-H. *Mean-Field Magnetohydrodynamics and Dynamo Theory* (Pergamon Press, Oxford/Akademie-Verlag, Berlin, 1980).
- Starchenko, S. V. & Shukurov, A. M. *Preprint No. 1310* (Space Res. Inst., Moscow, 1987) (submitted to *Astr. Astrophys.*).
- Stix, M. *Astr. Astrophys.* **42**, 85–89 (1975). Erratum **68**, 459 (1978).
- White, M. P. *Astr. Nachr.* **299**, 209–216 (1978).
- Ruzmaikin, A. A., Sokoloff, D. D. & Shukurov, A. M. *Astr. Astrophys.* **148**, 335–343 (1985).
- Krasheninnikova, Yu. S., Ruzmaikin, A. A., Sokoloff, D. D. & Shukurov, A. M. *Preprint No. 1393* (Space Res. Inst., Moscow, 1988) (submitted to *Astr. Astrophys.*).
- Clemens, D. P. *Astrophys. J.* **295**, 422–436 (1985).
- Vallée, J. P. *Astr. Astrophys.* **136**, 373–377 (1984).
- Syrovaty, S. I. in *Interstellar Gas Dynamics. Proc. IAU Symp. No. 39* (ed. Habing, H. J.) 192–193 (Reidel, Dordrecht, 1970).
- Michel, F. C. & Yahil, A. *Astrophys. J.* **179**, 771–780 (1973).
- Rees, M. J. *Q. J. R. astr. Soc.* **28**, 197–206 (1987).
- Kemp, J. C. *Publs astr. Soc. Pacif.* **94**, 627–633 (1982).

Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes

Suzanne L. Mansour, Kirk R. Thomas & Mario R. Capecchi

Howard Hughes Medical Institute, Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

Gene targeting—homologous recombination of DNA sequences residing in the chromosome with newly introduced DNA sequences—in mouse embryo-derived stem cells promises to provide a means to generate mice of any desired genotype. We describe a positive and negative selection procedure that enriches 2,000-fold for those cells that contain a targeted mutation. The procedure was applied to the isolation of *hprt*[−] and *int-2*[−] mutants, but it should be applicable to any gene.

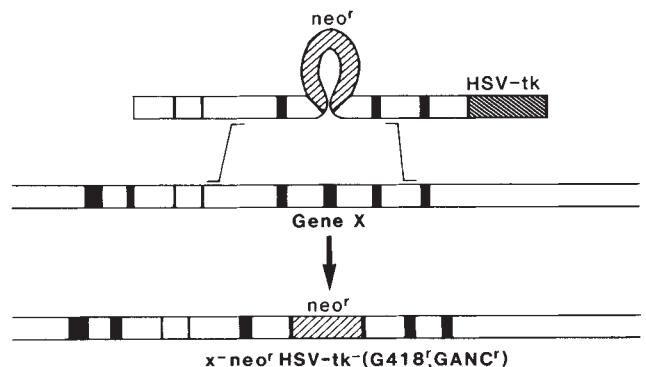
A GENERAL approach for producing mice of any desired genotype would involve the use of standard recombinant DNA techniques to introduce a desired mutation into a cloned DNA sequence of a chosen locus. That mutated sequence would then be transferred into an embryo-derived stem (ES) cell genome by gene targeting^{1,2}. Microinjection of the mutant ES cells into mouse blastocysts could then be used to generate germ-line chimaeras³, and finally, interbreeding of heterozygous siblings would yield animals homozygous for the desired mutation.

A major technical difficulty in this approach has been the unavailability of practical methods for obtaining ES cell lines carrying desired but non-selectable, targeted mutations at loci of interest. Previous studies have shown that we could specifically inactivate the *hprt* (hypoxanthine-guanine phosphoribosyl transferase) locus in murine ES cells¹. The *hprt* gene was chosen for these experiments because its location on the X chromosome meant that only one mutant copy was needed to yield a phenotype in male cells. Furthermore, *hprt*[−] ES cells could be directly selected by growth in the presence of the cytotoxic base analogue 6-thioguanine (6TG). The results from these experiments established that gene targeting in ES cells was feasible and defined some of the parameters that control the efficiency of this process. Most genes, however, are present as two copies in the genome, and a selectable cellular phenotype is not associated with the inactivation of the vast majority of these genes. Therefore, isolation of the rare ES cell in which a non-selectable gene has been inactivated by gene targeting must be accomplished by using a suitable enrichment and/or screening procedure.

Here we describe a general method for isolating ES cells containing targeted mutations in any gene, regardless of its function. A cloned fragment of the gene must be available and the intron-exon boundaries within that fragment defined. No other information is required. The procedure that we have developed uses a positive selection for cells that have incorporated the targeting vector anywhere in the ES genome (that is, either at the target site or at random sites) and a negative selection against the cells that have randomly integrated the vector into their genomes. The net effect is to enrich for cells containing the targeted mutation. This procedure was used to isolate ES cells containing insertions in the *hprt* and *int-2* genes.

The *int-2* proto-oncogene was first identified as a gene activated in mammary tumors of mice by the nearby insertion of the mouse mammary tumour virus^{4,5}. The *int-2* gene codes for a protein with significant similarity to the fibroblast growth factor family⁶, and its expression is highly restricted during mouse development. *int-2* messenger RNAs are detectable at very low levels in ES cells (less than one copy per cell) and are induced following *in vitro* differentiation of ES cells along the extra-embryonic endodermal lineage^{7,8}. *In situ* hybridization

a Gene Targeting



b Random Integration

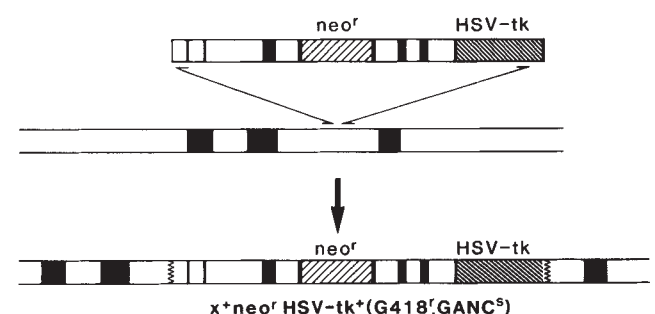


Fig. 1 The PNS procedure used to enrich for ES cells containing a targeted disruption of gene X. *a*, A gene X-replacement vector, that contains an insertion of the *neo*^r gene in an exon of gene X and a linked HSV-*tk* gene, is shown pairing with a chromosomal copy of gene X. Homologous recombination between the targeting vector and genomic X DNA results in the disruption of one copy of gene X and the loss of HSV-*tk* sequences. Such cells will be X[−], *neo*^r and HSV-*tk*[−] and will be resistant to both G418 and GANC. *b*, Because non-homologous recombination between the targeting vector and genomic DNA occurs through the ends of the linearized DNA^{9–11}, the HSV-*tk* gene remains linked to the *neo*^r gene. Such cells will be X⁺, *neo*^r and HSV-*tk*⁺ and therefore resistant to G418 but sensitive to GANC. Open boxes denote introns or flanking DNA sequences, closed boxes denote exons and cross-hatch boxes denote the *neo*^r or HSV-*tk* genes.

analysis of mouse embryo sections has shown that *int-2* is expressed specifically in the parietal branch of the endodermal lineage⁹. This analysis also showed that *int-2* transcripts are found in the embryo proper at 7.5–9.5 days of development: in mesodermal cells migrating through the primitive streak, in the neuroepithelium adjacent to the developing otocyst, and in the endoderm of the pharyngeal pouches. No *int-2* expression has been detected in the adult mouse. This diverse but highly restricted pattern of expression suggests that *int-2* may have several different functions during embryogenesis. To define these roles we have initiated a genetic analysis of *int-2* in the mouse.

Strategy

In experiments designed to introduce an exogenous DNA sequence at a specific, target locus in the genome, the vast majority of transfected cells contain the exogenous sequence inserted at random loci. Figure 1 outlines our experimental design to enrich for those rare cells in which the exogenous DNA recombined with its chromosomal cognate. The recombinant vector used for these experiments is of the replacement class¹. It contains 10–15 kilobases (kb) of DNA homologous to the target gene, *X*; a neomycin resistance (*neo*^r) gene inserted into an exon; and a Herpes simplex virus thymidine kinase gene (*HSV-tk*) adjacent to the target homology. The purpose of using the *neo*^r gene in the target vector is twofold: first, to disrupt the coding sequence of gene *X* and second, to act as a selectable marker (conferring resistance to the drug G418) for cells containing an integrated copy of the recombinant vector. To maximize the number of chromosomal integration sites that are capable of expressing *neo*^r, we have used pMC1Neo, a *neo*^r gene that was modified for efficient expression in ES cells¹. The *HSV-tk* gene was similarly placed under the influence of the promoter/enhancer that controls the *neo*^r gene.

The vector is designed so that when replacement of the endogenous *X* sequence by the exogenous one occurs via homologous recombination, the *HSV-tk* gene will not be transferred into the endogenous target (Fig. 1a). Exclusion of the *HSV-tk* gene during homologous recombination occurs because the *HSV-tk* gene represents a discontinuity in the incoming vector between homology and non-homology with the endogenous target sequence. Cell lines in which the targeting event occurred will therefore be *X*[−], *neo*^r, *HSV-tk*[−]. On the other hand, random integration of the target vector into the recipient cell genome should result in most cases, in cells that are *X*⁺, *neo*^r, *HSV-tk*⁺ (Fig. 1b) as we and others have shown that the majority of random insertions of exogenous, linearized DNA into the genome occurs through the ends^{10–12}. Therefore, by selecting for cells containing a functional *neo*^r gene (G418^r) and against cells containing a functional *HSV-tk* gene (gancyclovir resistance, GANC^r) we enrich for cells in which the targeting event had occurred.

As Herpes infections are a delicate human condition, researchers have actively sought nucleoside analogues that would selectively block viral propagation as a result of their specific metabolism by the viral thymidine kinase^{13–16}. Such drug design is possible because the substrate requirements of the *HSV-tk* protein are less stringent than those of the cellular *tk* enzyme. We tested two such nucleoside analogues, acyclovir and gancyclovir, for their cytotoxic effects on ES cells that contain a functional *HSV-tk* gene. Gancyclovir proved to be more effective. In Fig. 2, we compare the colony forming efficiency of ES cells with two *HSV-tk*⁺ ES cell lines as a function of gancyclovir concentration. It is apparent that over the concentration range tested, 10^{−8}–10^{−5} M, gancyclovir is not cytotoxic to the parental ES cell line but selectively kills ES cells containing *HSV-tk*.

Enriching for *hprt* targeting events

The feasibility of using the experimental design depicted in Fig. 1 to enrich for cells in which a targeting event has occurred was

first tested by using the endogenous *hprt* gene as the target sequence. Using this gene, we can directly compare the results obtained by direct selection for *hprt*[−] with those obtained by using the enrichment procedure.

In Fig. 3a we illustrate the *hprt* targeting vector used for these experiments. It represents the previously described *hprt* replacement vector pRV9.1 (ref. 1) to which we have added the *HSV-tk* gene at the 5' end. The linearized DNA was introduced into ES cells by electroporation and the cells were then subjected to different growth conditions (see Table 1 legend). The results of these experiments are summarized in Table 1. Of the cells, 50–60% survived electroporation. The number of G418^r, G418^r–6TG^r and G418^r–GANC^r colonies per 10⁶ survivors was approximately 10⁴, 5 and 4 respectively. DNA samples from six of the G418^r–6TG^r and 24 of the G418^r–GANC^r cell lines were analysed by Southern transfer. As expected, all of the G418^r–6TG^r cell lines were *hprt*[−] as a result of homologous recombination between the targeting vector and the endogenous *hprt* gene. None of these cell lines contained *HSV-tk* sequences. Thus, as intended, these sequences were eliminated from the incoming vector during the homologous recombination events. Of the 24 G418^r–GANC^r cell lines analysed, 19 contained the targeted mutations in the endogenous *hprt* gene. These cell lines were also 6TG^r and lacked *HSV-tk* sequences. Each of the original colonies was of independent origin (that is, obtained from separate electroporation experiments and on separate plates). From these data we can conclude that the enrichment procedure was successful: the majority of G418^r–GANC^r ES cell lines contained targeted disruptions of the *hprt* gene even though we did not directly select for the *hprt*[−] phenotype.

In Fig. 4 we compare the *hprt* Southern transfer pattern of the parental ES cell line with that of two cell lines electroporated with pRV9.1/TK and subjected to either G418 plus 6TG selection (line 40-9n), or G418 plus GANC selection (line 42-6j). The DNA from both of the latter two cell lines is altered at the

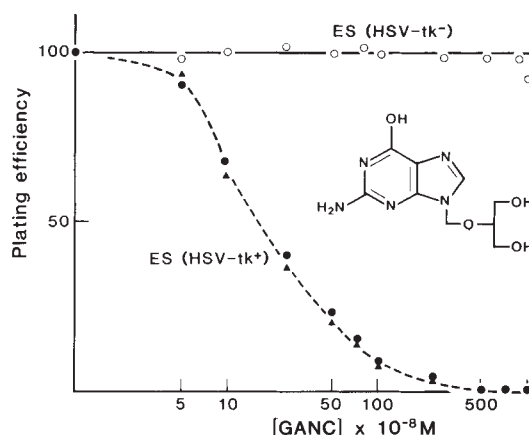


Fig. 2 The effect of gancyclovir on the plating efficiency of ES cells and two derivative, *HSV-tk* containing, ES cell lines. ES cells (2×10^4) (○) or two derivative ES cell lines (● or ▲), each containing a single copy of the *HSV-tk* gene, were plated on 100-mm petri plates containing Dulbecco's modified Eagle's medium, 10% fetal calf serum and the indicated concentration of GANC. The cells were incubated at 37°C in 7% CO₂. Every three days the cells were fed with fresh medium. After nine days, the plates were stained with Giemsa and the number of colonies per plate determined. The normalized number of colonies from the plates without GANC is represented by 100. Each point represents data obtained by averaging the results from three plates. *HSV-tk*⁺ cells were obtained by electroporating a plasmid containing the *neo*^r gene and the *HSV-tk* gene into the ES cells and selecting for G418^r cells. The *HSV-tk* gene contained the same promoter/enhancer present in the *neo*^r gene of pMC1Neo (ref. 1). The insert illustrates the structure of GANC.

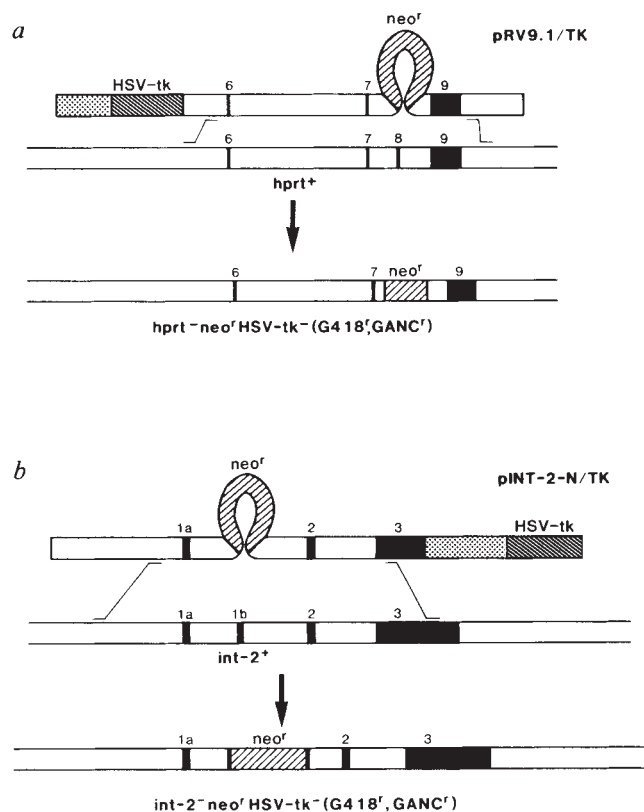


Fig. 3 Targeted disruption of the *hprt* locus (a) or the *int-2* locus (b) by the replacement vectors pRV9.1/TK or pINT-2-N/TK respectively. a, pRV9.1/TK contains 9.1 kb of DNA from the 3' end of the murine *hprt* gene, exons 6 to 9, with the eighth exon disrupted by the *neo^r* gene from pMC1Neo (ref. 1), and the *hprt-neo^r* sequences ligated to the HSV-tk gene. After homologous pairing between the linearized targeting vector and the genomic *hprt* gene, a recombination event replaces the endogenous *hprt* sequences with sequences that contain the *neo^r* insertion. The HSV-tk gene is not transferred to the genome. The cell line produced by this recombination event will be *hprt⁻*, *neo^r* and HSV-tk⁻ and will thus be resistant to both G418 and GANC. Stippled boxes denote plasmid sequences. The *int-2* replacement vector, pINT-2-N/TK was constructed starting with a 10-kb *int-2* genomic clone containing exons 1a, 1b, 2 and part of 3 in pAT153 (ref. 4). The *neo^r* gene from pMC1Neo was inserted at the *Apa*I site in exon 1b, the first protein-coding exon, disrupting it, and creating pINT-2-N. The HSV-tk gene, with the same enhancer-promoter as the *neo^r* gene of pMC1Neo, was inserted adjacent to pINT-2-N at its 3' end to create pINT-2-N/TK. Homologous recombination between the vector and genomic *int-2* sequences results in the disruption of one copy of the *int-2* gene and the loss of the HSV-tk gene. These cells are *int-2⁻/int-2⁺*, *neo^r* and HSV-tk⁻, and thus resistant to both G418 and GANC.

hprt locus as predicted for a targeting event. Digestion of parental ES DNA with *Bgl*II and *Eco*RI isolates sequences homologous to the *hprt* probe on fragments of 5.4 kb and 9.3 kb in length, respectively; the patterns of 40-9n and 42-6j are quite different, with fragments of 6.4 kb and 8.3 kb respectively. Because the *neo^r* gene has no *Bgl*II site, the size of the genomic *Bgl*II fragment is increased by the length of the *neo^r* insert (~1 kb). However, an *Eco*RI site near the 5' end of the *neo^r* gene reduces the genomic *Eco*RI fragment by ~1 kb (see Fig. 4b).

Targeting into the *int-2* gene

Figure 3b illustrates the vector used to target a mutation to the chromosomal *int-2* gene. This targeting vector, pINT-2-N/TK,

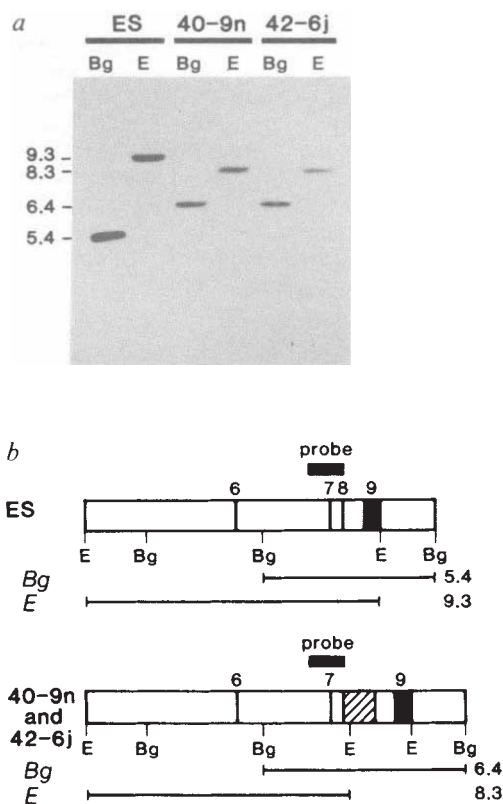


Fig. 4 Southern transfer demonstration of gene targeting by the targeting vector pRV9.1/TK. ES refers to the DNA from the parental, wild-type ES cell line, 40-9n refers to DNA from a transfected cell line resistant to G418 plus 6TG, and 42-6j refers to DNA from a transfected cell line selected for G418^r plus GANC^r. The length of the fragments is in kb. b, A schematic representation of the Southern transfer data. The top map represents the 3' end of the *hprt* gene from ES cells; the bottom map represents the *hprt* gene from a targeted cell line. Beneath each gene is shown the length and position of restriction fragments hybridizing to the probe. Bg, *Bgl*II; E, *Eco*RI. The very light hybridization signal observed in 40-9n and 42-6j at the parental position (that is, 5.4 kb and 9.3 kb) comes from the feeder cells upon which the ES cells are grown¹.

Methods. ES cells were transfected with pRV9.1/TK and selected for G418^r plus 6TG^r (line 40-9n) or G418^r plus GANC^r (line 42-6j) as described in the Table 1 legend. Following the isolation of individual cell lines, DNA was purified and digested with restriction endonuclease. DNA (7 µg) was electrophoresed through agarose, transferred to nitrocellulose and hybridized to a ³²P-labelled probe from the murine *hprt* gene¹.

was linearized as shown, introduced into ES cells by electroporation, and subjected to one of three growth conditions. The results from these experiments are summarized in Table 1. It is evident that G418 plus GANC selection resulted in a 2,000-fold enrichment for ES cells containing a disrupted *int-2* gene. Approximately one in 40,000 G418^r colonies contained an *int-2* mutation. Of the 81 independent G418^r-GANC^r cell lines tested, four (5%), contained a mutant *int-2* gene.

DNA from cells resistant to G418 plus GANC was subjected to Southern transfer analysis to assay for the presence of the targeted disruption in one copy of the *int-2* gene. In Fig. 5 we illustrate Southern transfer analysis of one of the G418^r-GANC^r cell lines, 43-8H, in which one of the two *int-2* autosomal genes was disrupted by the *neo^r* gene. Parental ES DNA that was digested with *Hind*III plus *Xho*I, *Bst*EII plus *Xho*I or *Xmn*I alone and hybridized with the 10-kb genomic *int-2* probe, shows the expected fragments of 11.2, 16 and 6.2 kb doublet respec-

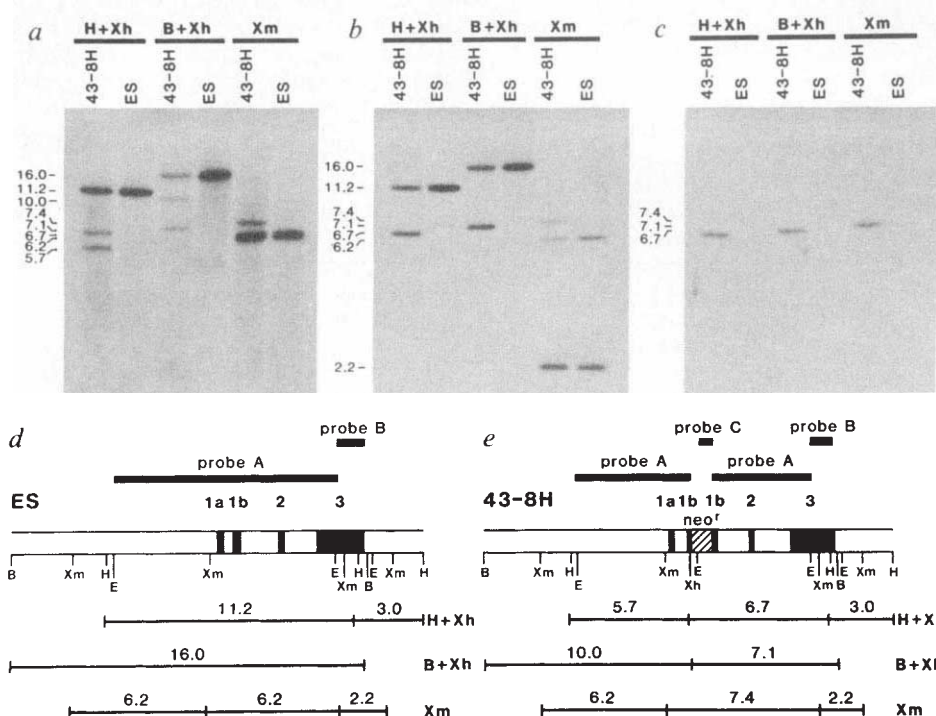


Fig. 5 Southern transfer analysis of DNA from ES and 43-8H cells. Top: the DNA samples and the enzymes with which they were digested are indicated above each lane. The membranes were hybridized with a 10-kb *EcoRI* fragment from *int-2* (Probe A, a), a 1.3 kb *int-2* flanking DNA fragment not present in the targeting vector (Probe B, b) or a 0.6-kb *PstI* to *BamHI* neo fragment from pMC1Neo (Probe C, c). a, Digestion of DNA from ES cell line 43-8H revealed new fragments compared with the digestion pattern of DNA from the parental ES cell line. Digestion with *HindIII* plus *XhoI* generates new fragments of 6.7 and 5.7 kb; *BstEII* plus *XhoI* digestion generates new fragments of 10 and 7.1 kb; and *XmnI* digestion generates a new fragment of 7.4 kb. b, As expected, probe B hybridizes with a *HindIII* plus *XhoI* fragment of 11.2 kb, a *BstEII* plus *XhoI* fragment of 16 kb, and *XmnI* fragments of 6.2 and 2.2 kb in both the parental and 43-8H cell lines. Additional DNA fragments of 6.7 kb (*HindIII* plus *XhoI*), 7.1 kb (*BstEII* plus *XhoI*) and 7.4 kb (*XmnI*) that hybridize with probe B were detected in 43-8H. d, e, Restriction maps of the parental ES DNA and the targeted DNA (43-8H) in the vicinity of *int-2* are depicted. Open boxes represent introns or flanking DNA, closed boxes indicate the positions of *int-2* exons 1a, 1b, 2 and 3. The *neo^r* gene is denoted with a cross-hatched box. The position of restriction sites used in the analysis are indicated with the letters B (*BstEII*), E (*EcoRI*), H (*HindIII*), Xh (*XhoI*) and Xm (*XmnI*). Above each map are boxes showing the map positions of the probe fragments used in the hybridization analysis and below each map is a schematic depiction of the expected fragments from each restriction enzyme digestion.

Table 1 Gene targeting into the *hprt* and *int-2* genes

Vector	Number of cells surviving electroporation	Number of G418 ^r colonies	Number of G418 ^r -6TG ^r colonies	Number of G418 ^r -GANC ^r colonies	Fraction of G418 ^r -GANC ^r colonies containing targeting events
pRV9.1/TK	1.3×10^7	1.5×10^5	64*	48	19/24
pINT-2-N/TK	1.5×10^7	1.6×10^5	—	81	4/81

Linearized DNA, as depicted in Fig. 3, was introduced into ES cells by electroporation. Aliquots of cells were then subjected to one of three (pINT-2-N/TK) or four (pRV9.1/TK) growth conditions: non-selective medium to assess the total number of cells surviving electroporation; G418 medium to assay the fraction of survivors transformed by the *neo^r* vector; G418 plus 6TG medium (pRV9.1/TK only) to select directly for cells simultaneously containing the *neo^r* gene but lacking a functional *hprt* gene; and G418 plus GANC medium to enrich for, but not directly select for, cells that contain a functional *neo^r* gene as a result of targeting into the *hprt* or *int-2* genes. ES cells were transfected by electroporation with $25 \mu\text{g ml}^{-1}$ of linearized pRV9.1/TK or pINT-2-N/TK. The conditions for electroporation, cell culture, selection for G418^r, G418^r-6TG^r or G418^r-GANC^r cell lines were as previously described¹ except that in G418-GANC medium, 6TG was substituted by 2×10^{-6} M GANC (gift of Syntex Research). Targeting events were identified by Southern transfer analysis (see Figs 4 and 5).

*To evaluate the number of G418^r-6TG^r colonies, 1/8 of the cells surviving electroporation were grown in G418, 6TG medium. We obtained eight independent G418^r-6TG^r colonies. The value shown (8×8) was normalized to the total number of cells surviving electroporation.

tively (Fig. 5a). DNA from ES cell line 43-8H that was digested with the same enzymes, reveals fragments the same size as those from the parental line (that is, derived from the wild-type copy of *int-2*), but also reveals new fragments derived from the mutagenized copy of *int-2* (Fig. 5a). Each of these additional fragments is the size predicted from a homologous recombination event between the targeting vector and one copy of the

wild-type *int-2* gene (see Fig. 5d, e and legend). In addition, these new bands are of the same intensity as those derived from the unmutagenized chromosome, suggesting that the cell line is a pure population, heterozygous at the *int-2* locus.

The structure of the mutant allele was further confirmed by hybridizing separate copies of the blot in Fig. 5a with two probes: B, a 3' flanking *int-2* DNA fragment not present in the

targeting vector (Fig. 5b) and C, a DNA fragment derived from the *neo^r* gene (Fig. 5c). Hybridization with probe B revealed additional fragments in the digest of the 43-8H cell line not present in the parental one (Fig. 5b and legend). This result is only consistent with an event in which the DNA from the *neo^r*-containing targeting vector (which has no sequence homology to probe B) has replaced one copy of the wild-type *int-2* gene. Finally, the *neo^r* DNA fragment (probe C) hybridizes only with DNA from 43-8H cells, and the hybridizing fragments align precisely with the unique fragments that were also detected with probe B (Fig. 5c).

In summary, the map positions of 10 restriction sites that cover approximately 20 kb of genomic DNA surrounding the *neo^r* insertion in the *int-2* gene of 43-8H were confirmed by Southern transfer analysis. Several additional enzymes, including *EcoRI* which delimits the ends of the *int-2* sequences in the target vector, have been used to extend the above results (data not shown). In none of the four *int-2⁻/int-2⁺* cell lines have we detected any deviations from the pattern predicted for a homologous replacement of chromosomal *int-2* sequences with DNA from the targeting vector. This suggests that the targeting event did not induce any gross rearrangements of the target DNA other than the *neo^r* insertion.

The (*int-2⁻/int-2⁺*) ES cell lines that we have isolated are capable of differentiating *in vitro* and of efficiently generating chimaeric mice following their introduction into recipient blastocysts. We are therefore encouraged that they will be capable of forming germ line chimaeras. The offspring of such chimaeras could then be interbred to assess the phenotypic consequences of the homozygous *int-2⁻* condition.

Discussion

We have described an experimental design for isolating ES cells that contain targeted modifications of any endogenous gene, irrespective of the function of that gene. This procedure uses a combination of a positive selection, growth in the presence of G418, for cells containing the input recombinant vector, and a negative selection, growth in the presence of GANC, against cells containing random integrations of the incoming recombinant vector. The net effect of these two selections is to enrich for cells in which the targeted homologous recombination event has occurred. The procedure has been applied to the isolation of cell lines that contain targeted disruptions of the *hprt* and *int-2* genes. In each case, the positive-negative selection (PNS) procedure resulted in a 2,000-fold enrichment of cells that contained homologous versus non-homologous integrations of the targeting vector. In the case of *hprt*, as the ratio of homologous events to random integration events was approximately one in two thousand, nearly every G418^r-GANC^r colony that survived PNS treatment (19/24) contained a disrupted *hprt* gene.

Using a similar targeting vector, the relative and absolute frequencies of homologous recombination at the *int-2* locus were 20-fold lower than those at the *hprt* locus. This lower targeting frequency may reflect the lower level of *int-2* expression in ES cells (that is, less than one transcript per cell)⁷⁻⁹. The level of expression and the chromatin environment surrounding the target gene could influence the targeting efficiency at that

locus by influencing accessibility to the recombination machinery and/or by influencing the expression of the *neo^r* gene, within that environment. Arguing against the second point, all of the ES cell lines that contain mutant *int-2* genes are robust in G418 medium, even though *int-2* expression is barely detectable in ES cells.

At this stage it is difficult to predict how much the targeting frequency will vary from locus to locus. However, for genes that are recalcitrant to targeted disruption, the enrichment factor for the PNS procedure may be increased from 2×10^3 to 4×10^6 by inserting into the targeting vector two HSV-*tk* genes, one at each end. The reason that a higher enrichment was not obtained with the single HSV-*tk* constructs is that some of the HSV-*tk* genes are mutated during transfection (unpublished results). In the PNS procedure we demand that the cells acquire a functional *neo^r* gene. By placing two HSV-*tk* genes at opposite ends of the target vector, it becomes more likely that one will survive the transfection procedure unimpaired. If the frequency of one HSV-*tk* acquiring a deleterious mutation during transfection is one in 2×10^3 molecules, then the frequency of inactivating both HSV-*tk* genes should be one in 4×10^6 molecules.

It was somewhat surprising to find that placing a large block of non-homology, such as an HSV-*tk* gene, at both ends of the linearized targeting vector does not reduce its ability to participate in homologous recombination with the cognate chromosomal gene (K.R.T. and M.R.C., unpublished results). This observation at first appears paradoxical. On the one hand, linearized vectors are much better targeting substrates¹². On the other hand, blocking both ends of the linearized vector with non-homologous DNA does not reduce its targeting efficiency. These results indicate that linearizing the target vector does not increase its targeting efficiency by providing homologous ends for invasion of the target gene. Rather, linearization may convert the targeting vector into a better topological substrate for participation in homologous recombination with the chromosomal gene.

Conclusion

We have demonstrated that the PNS procedure effectively enriched for ES cells containing targeted disruptions of either the *hprt* or *int-2* locus. The same procedure has also been successfully applied to isolate ES cells containing a mutant mouse homoeobox gene, *hox1.2* (D. Kostic and M.R.C., unpublished results). This procedure is general and requires little knowledge of the target locus. We believe that the procedure should be applicable to any gene. It has worked for a gene that is expressed in ES cells at very low levels, *int-2*, but it remains to be determined whether the procedure is applicable to genes with no detectable expression in ES cells.

We would like to acknowledge the technical assistance of Susan Tamowski, Carol Lenz and Laurie Fraser. We would also like to thank Drs Ron Evans, Salk Institute and Julian Verheyden, Syntex Research for suggesting to us that gancyclovir might be better than acyclovir as a cytotoxic agent against HSV-*tk* containing cells. S.L.M. and K.R.T. were supported by fellowships from the American Cancer Society and Cystic Fibrosis Foundation respectively.

Received 15 September; accepted 14 October 1988.

1. Thomas, K. R. & Capecchi, M. R. *Cell* **51**, 503-512 (1987).
2. Doetschman, T. *et al.* *Nature* **330**, 576-578 (1987).
3. Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. *Nature* **309**, 255-256 (1984).
4. Peters, G., Brookes, S., Smith, R. & Dickson, C. *Cell* **33**, 369-377 (1983).
5. Dickson, C., Smith, R., Brookes, S. & Peters, G. *Cell* **37**, 529-536 (1984).
6. Dickson, C. & Peters, G. *Nature* **326**, 833 (1987).
7. Jakobovits, A., Shackelford, G. M., Varmus, H. E. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7806-7810 (1986).
8. Smith, R., Peters, G. & Dickson, C. *EMBO J.* **7**, 1013-1022 (1988).

9. Wilkinson, D. G., Peters, G., Dickson, C. & McMahon, A. P. *EMBO J.* **7**, 691-695 (1988).
10. Folger, K. R., Wong, E. A., Wahl, G. & Capecchi, M. R. *Molec. cell. Biol.* **2**, 1372-1387 (1982).
11. Roth, D. B., Porter, T. N. & Wilson, J. H. *Molec. cell. Biol.* **5**, 2599-2607 (1985).
12. Thomas, K. R., Folger, K. R. & Capecchi, M. R. *Cell* **44**, 419-428 (1986).
13. St. Clair, M. H., Miller, W. H., Miller, R. L., Lambe, C. U. & Furman, A. *Antimicrobial Agents and Chemotherapy* **25**, 191-194 (1984).
14. Chen, M. S., Amico, L. A. & Speelman, D. J. *Antimicrobial Agents and Chemotherapy* **26**, 778-780 (1984).
15. McLaren, C., Chen, M. S., Barbhuiya, R. H., Burober, R. A. & Olsen, F. B. in *Herpes Virus and Virus Chemotherapy* (ed. Kono, R.) 57-61 (Elsevier, Amsterdam, 1985).
16. Mansuri, M. M. *et al. J. Med. Chem.* **30**, 867-871 (1987).

Interaction at a distance between λ repressors disrupts gene activation

A. Hochschild & M. Ptashne

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

The λ repressor is an activator as well as a repressor of transcription. The activation function is blocked by interaction with another λ repressor molecule bound upstream on the same DNA molecule. This example of negative control at a distance involves formation of a DNA loop.

ONE of the salient features of gene regulation in both prokaryotes and eukaryotes is action at a distance: the repression and activation of transcription by regulatory proteins bound to DNA far from the startpoint of transcription. According to one prevalent view, action at a distance involves the interaction between proteins bound to separated sites, the DNA in between looping out to accommodate the interaction (for review, see ref. 1). Such a reaction has been demonstrated, for example, using purified λ repressor^{2,3}; this protein binds cooperatively not only to adjacent operator sites⁴, but also to sites that are separated by an integral number of turns of the DNA helix. Interaction between bound repressors, whether adjacent or separated, can be used to ensure repressor binding to a site that overlaps the promoter; such a strategy is evidently used to maintain efficient repression in several bacterial operons⁵⁻¹¹. Using λ repressor, we demonstrate another way in which negative regulation can work at a distance. That is, we show that an activator positioned adjacent to the promoter can interact with an identical molecule bound at a distance upstream to block gene activation. A similar mechanism has been proposed to explain negative regulation at a distance in the arabinose operon^{12,13}.

Figure 1 summarizes the ordinary activities of λ repressor as both a negative and a positive regulator of transcription. Repressor normally binds cooperatively to adjacent operator sites O_R1 and O_R2 , and blocks transcription of the early lytic genes when bound at these sites (for reviews, see refs 14 and 15). But repressor also turns on transcription of its own gene by a mechanism that we believe involves a contact between repressor bound at O_R2 and RNA polymerase bound at the adjacent promoter (P_{RM}) (refs 16-19). The only role of O_R1 (a strong repressor-binding site) is to help ensure the occupancy of O_R2 (a weak binding site), and so at higher concentrations repressor can bind to O_R2 and activate transcription in the absence of repressor bound at O_R1 (ref. 16). Here we show that under conditions where repressor is bound at O_R2 , the binding of a second repressor six or seven helical turns upstream blocks activation. We show both *in vivo* and *in vitro* that this inhibition involves interaction between the separated repressors with the formation of a DNA loop. We take advantage of the effect to isolate a mutant repressor that does not interfere with transcriptional activation when bound an integral number of turns upstream, and show that it does not bind cooperatively either to separated or to adjacent operator sites *in vitro*.

Inhibition of activation *in vivo*

We generated a set of DNA templates bearing O_R1 at various distances upstream of O_R2 , as well as a template from which O_R1 was deleted (see legend to Fig. 2). Each template bears P_{RM} positioned at its normal site adjacent to O_R2 and fused to the *lacZ* gene. We show the effects of various concentrations of repressor, encoded by an added plasmid, on transcription from P_{RM} in cells bearing one or another of these templates.

Figure 2 shows the effect of repressor present at a concentration (hereafter called 'moderate') sufficient to occupy O_R2 in

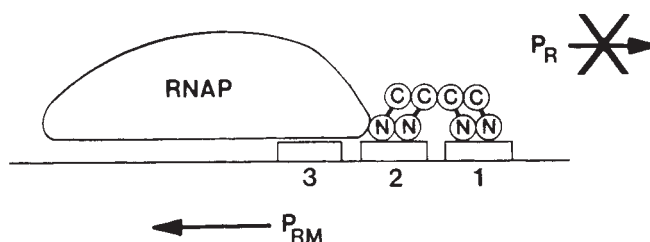


Fig. 1 Stimulation of transcription from P_{RM} in a λ lysogen. Repressor dimers are bound cooperatively to O_R1 and O_R2 , the dimer at O_R1 helping a second to bind to O_R2 ; the latter contacts RNA polymerase to activate transcription from P_{RM} . Repressor is folded into two domains²⁷: the amino domain contacting DNA and RNA polymerase^{28,29} and the carboxyl domain which facilitates dimer formation²⁷ and mediates interaction between dimers^{2,4}.

the absence of the helping effect of a repressor bound at O_R1 . Repressor efficiently stimulated transcription when O_R1 and O_R2 were separated by a nonintegral number of helical turns (5.5, 6.5 or 7.4), but surprisingly, not when the spacing was integral (5.9, 6.1 or 7.0 turns). This result suggests that repressors bound to the same side of the DNA helix interact, with the intervening DNA forming a loop, and that this interaction prevents repressor bound at O_R2 from stimulating transcription efficiently.

Two additional experiments performed *in vivo* support this interpretation. First, a repressor fragment lacking the carboxyl domain stimulated transcription efficiently, regardless of the spacing between O_R1 and O_R2 ; this is expected, because the amino domain alone binds DNA noncooperatively⁴. Second, repressor at higher concentrations caused less inhibition than repressor at moderate concentrations; repressor at these high concentrations tetramerizes in solution (ref. 20 and R. T. Sauer, unpublished results), and we imagine that the interaction of a bound repressor with a free repressor (forming a tetramer) competes with loop formation. The data supporting these results are shown in Table 1; a more complete description of the experiments is given in the legend.

Inhibition of activation *in vitro*

Figure 3a shows that purified repressor stimulates transcription from P_{RM} *in vitro* when O_R1 and O_R2 are separated by a nonintegral number (5.5) of turns (compare lanes 3 and 4), but has no effect when the sites are separated by an integral number (5.9) of turns (compare lanes 1 and 2). Figure 3a also shows that in both cases repressor turns off transcription from the divergent P_R promoter also present on these templates, as expected.

The DNase I footprint experiment of Fig. 3b, performed under the conditions of the transcription experiment, shows that O_R1

Fig. 2 Stimulation by repressor as a function of the spacing between O_R1 and O_R2 . The spacing between O_R1 and O_R2 was varied by inserting DNA into a restriction site located between O_R1 and O_R2 . The relationship between O_R2 and P_R remains fixed in all constructions, and P_{RM} was fused to the *lacZ* gene. In addition, operator site O_R3 (which overlaps P_{RM} , see Fig. 1) was inactivated by a quadruple mutation. Each template was introduced in single copy into the bacterial chromosome to generate six different strains. β -Galactosidase assays determined the level of transcription from P_{RM} ; the stimulation by repressor is plotted as a function of the spacing between O_R1 and O_R2 . Repressor in moderate concentrations (see text) was supplied by the plasmid pKB158 (ref. 26). To assess the basal level of transcription for each template, the strains were separately transformed with a plasmid that does not encode a repressor (pBR322). The fold stimulation by repressor was calculated for each strain by comparing basal and stimulated levels of transcription. The basal levels of β -galactosidase for the six strains were similar, ranging between 2.1 and 2.45 units. All assays were performed in duplicate with values differing by <5%.

Methods. The parental plasmid pEM90_RP was used to generate the set of derivatives with different spacings between O_R1 and O_R2 . pEM90_RP bears the O_R region of λ and the first 100 base pairs of the *cl* gene inserted into the *EcoRI*-*HindIII* backbone of pEMBL9 (ref. 30). The O_R region on this plasmid carries four mutations in O_R3 , three of which were introduced by site-directed mutagenesis and resulted in the creation of a *PvuII* restriction site¹⁹. The sequence of the mutant O_R3 site is 5'TACAGCTGCAAGGATA 3'. To generate the 5.5-turn construct, pEM90_RP was digested at the unique *HincII* site at the edge of O_R2 and a 34-base pair polylinker fragment from pEMBL8 (ref. 30) was inserted. Using the 5.5-turn construct as a parent, we generated all the other spacings between O_R1 and O_R2 , taking advantage of unique restriction sites carried on the polylinker fragment. Each construct was eventually transferred to another vector, pRW283 (ref. 31), which was designed to facilitate transfer to a phage for introduction into the bacterial chromosome (see below). The resulting plasmids each encode a fusion of the first 33 amino acids of repressor and the α -complementing fragment of β -galactosidase under the control of λ 's P_{RM} . A detailed description of all plasmid constructions is available on request. The plasmid-borne constructions were transferred to lysogenic phages using recombination *in vivo*. The recipient phage was a derivative of λ RZ11 (ref. 32) which bears the immunity region of phage i21. Phage i21RZ11 carries a *lacZ* gene that is lacking both promoter and ATG and therefore forms white plaques on a *lacZ*⁻ strain on X-Gal indicator plates. Each of the derivatives of pRW283 was recombined *in vivo* i21RZ11. The appropriate double recombination event³¹ results in the transfer of the O_R region to the phage so that P_{RM} directs the synthesis of a fusion of the first 33 amino acids of λ repressor to β -galactosidase. The phage form blue plaques on X-Gal indicator plates. Each recombinant phage was stably integrated into the *lacZ*⁻ host NK5031 (F'*lacZ*ΔMM5265suIII⁺NaI^R) as an i21 prophage.

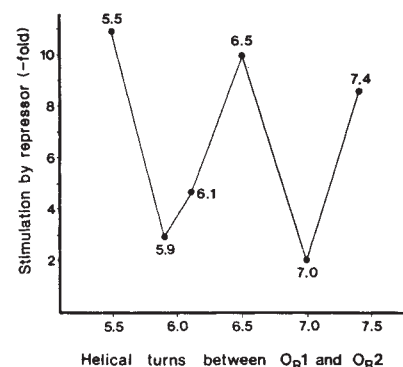


Fig. 3 Effect of repressor on transcription from P_{RM} *in vitro*. *a*, A run-off transcription experiment. Lanes 1 and 2 show the transcription products from the 5.9-turn template transcribed in the absence and presence of repressor and lanes 3 and 4 show the transcription products from the 5.5-turn template. The P_{RM} transcripts are different lengths because the two templates were truncated with different restriction enzymes (see below). Both templates carry a P_R promoter that has been damaged but not totally inactivated by the insertion of DNA between O_R1 and O_R2 . Transcription from the weakened P_R is repressed by the binding of repressor to O_R1 on these templates. The transcript visible at the bottom of the autoradiograph originates from another template (carrying a 434 promoter) that is included in the reactions as an internal control. *b*, DNase I protection experiment performed under the conditions of the transcription experiment. Lanes 1 and 2 show the digestion pattern of the 5.9-turn template digested in the absence and presence of repressor, and lanes 3 and 4 the digestion pattern of the 5.5-turn template. The concentration of repressor was the same as in the transcription reactions.

Methods. The run-off transcription assays were performed in 40 mM Tris, pH 7.9, 10 mM MgCl₂, 5 mM dithiothreitol, 125 μ g ml⁻¹ BSA, 10% (v/v) glycerol and 50 mM KCl in a total volume of 25 μ l. The DNA fragments (~10 nM), were incubated with or without repressor (at a concentration of 5.8×10^{-7} M) for 15 min. RNA polymerase (1 unit; New England Biolabs) was then added and the mixture incubated for another 10 min. Transcription was initiated by addition of 200 μ M ATP, 5 μ M GTP, 5 μ M CTP and ~6 μ M [α -³²P]UTP and heparin (100 μ g ml⁻¹). Transcription was stopped after 20 min at 37 °C by addition of 50 μ l phenol, ethanol-precipitated in 0.3M sodium acetate and resuspended in formamide dye for loading on to a 6% polyacrylamide/7M urea gel. The templates used were the *Bgl*II-*Bst*NI fragment of 2830_RP5.9 and the *Bgl*II-*Hind*III fragment of pEM90_RP5.5. The DNase I protection experiment was as described⁴, except that we used the transcription experiment conditions. To ensure that the DNA concentration was the same in both the footprint and transcription reactions, we used the same template mixtures in both, except that in the footprint experiments labelled DNA fragments were added at one tenth of the concentration of the unlabelled fragments. The concentrations of repressor when present were also identical in both reactions. No RNA polymerase or nucleotides were added to the footprint reactions. Both the *Bgl*II-*Bst*NI fragment carrying the 5.9-turn construct and the *Bgl*II-*Hind*III fragment carrying the 5.5-turn construct were 3' end-labelled at their *Bgl*II ends using [α -³²P]dGTP and reverse transcriptase.

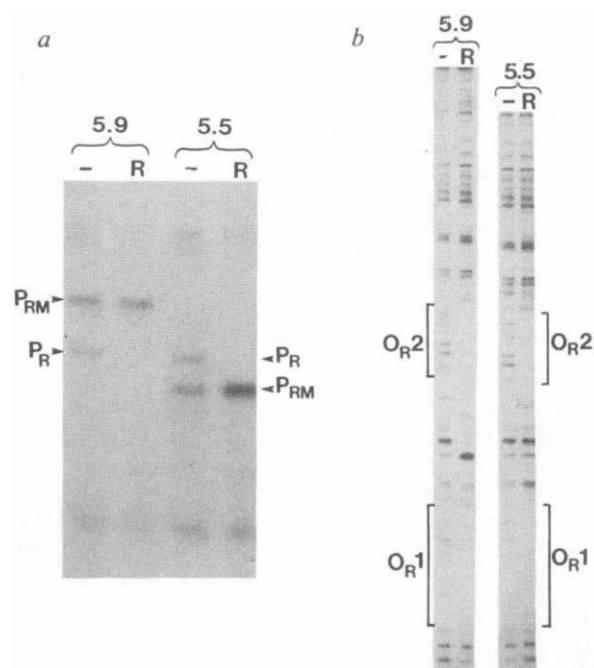
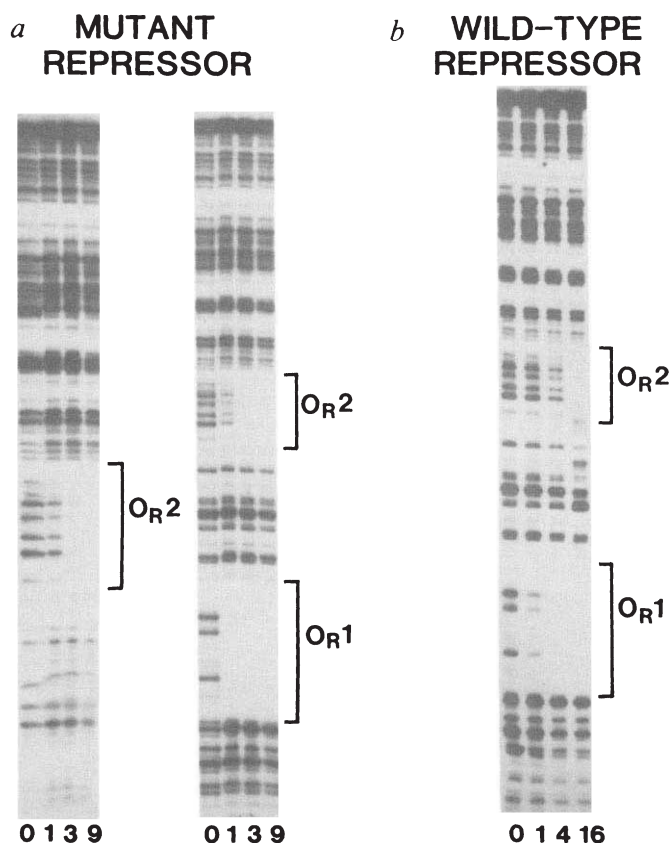


Fig. 4 Binding of mutant repressor to operator sites separated by 5.9 turns of the DNA helix. *a*, DNase protection experiment comparing the binding of the mutant repressor to O_{R2} in the absence or presence of O_{R1} 5.9 turns away on the same DNA fragment. Reactions were incubated with increasing concentrations of protein, a concentration of 1 corresponding to 1.5 μ M repressor monomer. There are no enhanced or diminished cleavages between the two operators on the 5.9-turn template and the same concentration of repressor is required to fill O_{R2} on both templates, indicating that the binding is noncooperative. *b*, Binding of wild-type repressor to the 5.9-turn template. A repressor concentration of 1 corresponds to 0.8 nM repressor monomer. The enhanced and diminished cleavages indicative of DNA loop formation are visible in the region between O_{R1} and O_{R2} .

Methods. Binding reactions were essentially as described⁴, except that the KC1 concentration was 20 mM and incubation was at 0 °C. These conditions were chosen because repressor binds more tightly at low salt and low temperature³³ and our preparation of mutant protein was relatively inactive. The two restriction fragments used were the *Bgl*III–*Bst*NI fragment of 2830_RP5.9 and the *Eco*RI–*Bst*NI fragment of 2830_RPΔ O_{R1} . The *Bgl*III–*Bst*NI fragment was labelled as described in Fig. 3 legend and the *Eco*RI–*Bst*NI fragment was 3' end-labelled at the *Eco*RI end using [α -³²P]dATP and reverse transcriptase. The plasmid 280Δ-35 was constructed for mutagenesis of the repressor gene. Although pKB158 directs the synthesis of appropriate (moderate) levels of repressor, we did not want to mutagenize this plasmid because expression of the repressor gene is under autogenous control and therefore a mutant repressor unable to bind cooperatively would probably be under-expressed. We therefore constructed a lower-expressing derivative of pKB280 (ref. 26), a plasmid that directs the synthesis of very high levels of repressor under the control of the *lacUV5* promoter. We used a *Hpa*II restriction site located between the –10 and the –35 regions of the *lac* promoter on pKB280 and purified a *Hpa*II–*Cla* fragment carrying the –10 region and the entire *cI* gene. After filling out the *Hpa*II end of this fragment, we inserted it into the *Eco*RI–*Cla* backbone of pBR322 with a filled-out *Eco*RI end. The resulting plasmid (280Δ-35) lacks the –35 region of the *lac* promoter and directs the synthesis of levels of repressor comparable with pKB158. The repressor mutation was first mapped to the carboxy-terminal third of the *cI* gene using the *Hind*III site located in the gene. A second *Hind*III site in 280Δ-35 is located 3' to the repressor gene and so we purified the *Hind*III fragment from the mutant plasmid and re-cloned it into an unmutagenized backbone. We then sequenced³⁴ the mutation by purifying a 455-base pair *Hind*III–*Pst*I fragment and 3' end-labelling the *Hind*III end using [α -³²P]dATP and reverse transcriptase. The C to T transition at nucleotide 686 on the noncoding strand was confirmed by 3' end-labelling a *Hinf*I site located 3' to the gene and sequencing the other strand. To purify the mutant repressor protein we first transferred the mutant gene to a *tac* vector³⁵. To this end, we purified a *Hind*III–*Cla* fragment (with a filled-out *Cla* end) from 280Δ-35ser→asn and inserted it into the *Hind*III–*Bam*HI backbone (with a filled-out *Bam*HI end) of pEA305 (ref. 35), a plasmid that directs the synthesis of very high levels of λ repressor under the control of the *tac* promoter. We thus replaced the carboxy-terminal portion of the wild-type repressor gene on pEA305 with that of the mutant gene. Purification³³ of the final protein preparation was estimated as >90% by inspection of Vesterburg-stained Laemmli gels. The activity was not determined, but on the basis of the behaviour of the mutant protein *in vivo*, our preparation is of relatively low activity.



and O_{R2} are completely occupied by repressor on each template; this indicates that the failure of repressor to stimulate transcription from P_{RM} on the 5.9-turn template is not due to a failure to occupy O_{R2} . Also, in the case of the 5.9-turn template, the footprint in the presence of bound repressor shows an alternating set of enhanced and diminished DNase I cleavages between the two operators. This pattern is diagnostic of DNA-loop formation: the distortion of the DNA backbone that is induced by cooperative binding results in a change in the sensitivity to DNase I of the looped DNA, with enhanced cleavages mapping to the outside and diminished cleavages to the inside of the protein-DNA complex².

A cooperativity mutant

These findings suggested a simple screen for repressor mutants that are unable to interact when bound to separated operator sites; such a mutant should activate transcription equally well whether O_{R1} and O_{R2} are separated by an integral or a nonintegral number of turns. A strain harbouring the 5.9-turn construct forms pale blue colonies on the appropriate indicator plate in the presence of repressor, whereas a strain harbouring the 5.5-turn construct forms darker blue colonies. We anticipated that

the desired mutant would cause the former strain to form darker blue colonies indistinguishable from those formed by the strain carrying the 5.5-turn construct. We used a plasmid that directs the synthesis of sufficient repressor to fill O_{R2} , independently, and mutagenized it by passage through the strain mut D. We transformed a strain carrying the 5.9-turn construct with the mutagenized plasmid DNA, obtained dark blue colonies at a frequency of ~0.2%, and recovered plasmid DNA from one of them for further characterization.

We mapped the mutation to the carboxy-terminal portion of the *cI* gene by purifying the relevant restriction fragment and re-cloning it into an unmutagenized vector. DNA sequencing of this portion of the gene revealed a single G.C to A.T transition which results in the replacement of Ser 228 with Asn.

Unlike wild-type repressor, the mutant stimulates transcription *in vivo* equally well on templates bearing O_{R1} and O_{R2} separated by an integral and a nonintegral number of turns: there is an 11.8-fold increase on the 5.5-turn template and a 13.6-fold increase on the 5.9-turn template. This observation suggests that the mutant repressor binds independently to O_{R1} and O_{R2} on both templates. This was confirmed by the experiment in Fig. 4, which shows that the purified mutant protein, unlike wild-type repressor, binds noncooperatively to O_{R1} and

Table 1 Stimulation by intact repressor and by repressor's amino domain

Helical turns	Amino domain	Intact repressor	
		Moderate	High
5.5	9.5 ×	10.8 ×	7.9 ×
5.9	7.7 ×	2.9 ×	5.5 ×
6.1	6.4 ×	4.5 ×	5.2 ×
6.5	6.8 ×	10.0 ×	5.9 ×
7.0	5.5 ×	2.2 ×	3.6 ×
7.4	7.7 ×	8.5 ×	6.1 ×
Wild type	3.0 ×	5.3 ×	3.0 ×
ΔO _{R1}	4.2 ×	4.7 ×	3.9 ×

Each template (column 1) was integrated into the bacterial chromosome in single copy (see Fig. 2 legend). A horizontal comparison of the entries for each template reveals one pattern of response characteristic of the integral spacings and another characteristic of the nonintegral spacings. Transcription from the integral templates is most efficiently stimulated by the amino domain, least efficiently stimulated by moderate concentrations of intact repressor, and stimulated to an intermediate degree by high concentrations of repressor (lines 2, 3 and 5). Transcription from the nonintegral templates is most efficiently stimulated by moderate concentrations of repressor, rather less efficiently by the amino domain (which dimerizes very poorly and therefore does not bind as effectively as intact repressor), and even less efficiently by high concentrations of repressor (lines 1, 4 and 6). We do not know the reason for the modest inhibition of stimulation at high concentrations of repressor, but it may reflect some binding to the quadruply mutated O_{R3} (see Fig. 2 legend). Another possibility is that a tetramer bound at O_{R2} is impaired relative to the dimer with respect to stimulation. Lines 7 and 8 show the stimulatory effects of intact repressor and the amino domain on a wild-type template and on a template deleted for O_{R1}. Like the nonintegral but unlike the integral templates, these templates are most efficiently stimulated by moderate concentrations of repressor. The extent of the stimulation was less for these two templates than for the three nonintegral templates. Presumably this low stimulation reflects the higher basal level of these two templates, as all five templates direct the synthesis of a comparable level of β -galactosidase when maximally stimulated. β -Galactosidase assays were performed in duplicate (values differed by <5%). Repressor's amino domain was supplied at a high concentration by the plasmid pKB280ΔH3, a deletion derivative of pKB280 (ref. 26) which lacks the carboxyl third of the repressor gene. Repressor at a moderate concentration (~5 lysogen's-worth) was supplied by pKB158 (ref. 26) and repressor at a high concentration (~30 lysogen's-worth) by pKB252 (ref. 26). The extent of stimulation was determined by comparing the β -galactosidase level in cells containing a plasmid that does not encode repressor (pBR322) and in cells containing one of the protein-producing plasmids. The basal (unstimulated) levels of β -galactosidase measured for the integral and nonintegral templates were all similar at ~2.2 units (see Fig. 2 legend), but were higher for the wild-type template and the template deleted for O_{R1} at 6.1 and 4.1 units.

O_{R2} when they are separated by 5.9 turns. That is, repressor binds O_{R2} at the same concentration whether or not O_{R1} is present on the same DNA template. As expected, the mutant protein is unlike wild-type in that it does not produce the altered pattern of DNase I reactivity in the region between the two operator sites, which is consistent with its predicted inability to form loops (compare lane 8 of Fig. 4a with lane 4 of Fig. 4b).

Figure 5a shows that the mutant protein also binds noncooperatively to O_{R1} and O_{R2} when they are adjacent to each other as they are on the λ chromosome. For comparison we also show a protection experiment using the wild-type protein (Fig. 5b).

Discussion

Our experiments show that a distally bound λ repressor can prevent another repressor bound next to a promoter from activating transcription. Three lines of evidence argue that this effect

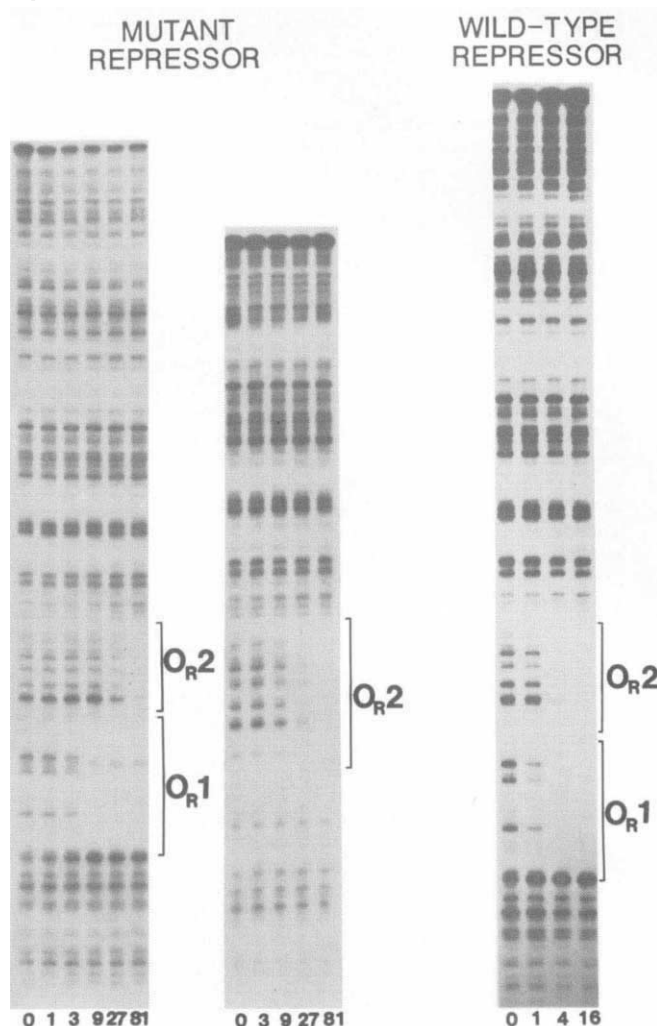


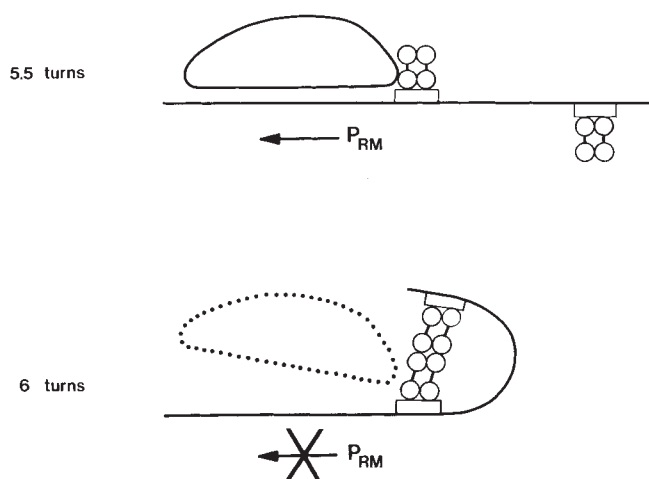
Fig. 5 Binding of mutant repressor to adjacent operator sites. *a*, DNase protection experiment comparing the binding of the mutant repressor to O_{R2} in the presence or absence of O_{R1} adjacent on the same DNA fragment. Reactions were incubated with increasing concentrations of protein, a concentration of 1 corresponding to 167 nM repressor monomer. Binding to the adjacent operators is noncooperative. *b*, Binding of wild-type repressor to adjacent operators. A repressor concentration of 1 corresponds to 0.8 nM repressor monomer. It takes <4 times as much repressor to fill O_{R2} as O_{R1}.

Methods. The reactions were performed at a KCl concentration of 20 mM and incubated at 0°C. The two restriction fragments used were the *Bgl*III-*Hind*III fragment of pFB80 and the *Eco*RI-*Bst*NI fragment of 2830_RΔO_{R1}.

involves an interaction between the bound repressor molecules with the formation of a DNA loop. First, the effect depends on the spacing between the two operator sites; the inhibition was observed only when the upstream operator was positioned an integral number of turns of the DNA helix away from the downstream operator (Fig. 6). Second, we replicated the effect *in vitro*, and showed that a DNase I-footprinting pattern that is diagnostic of DNA looping correlated with the inhibition of transcriptional activation. Finally, we isolated a mutant repressor that efficiently stimulated transcription in a strain carrying O_{R1} an integral number of turns upstream of O_{R2}; this mutant repressor binds noncooperatively to separated operator sites *in vitro*.

The mutant repressor that bears a single amino-acid change

Fig. 6 Model depicting inhibition of repressor-stimulated transcription. Repressors are shown bound independently to operator sites separated by a nonintegral number of turns (top). The repressor bound at O_{R2} contacts RNA polymerase bound at P_{RM} to stimulate transcription. Interaction is shown between repressors bound to operator sites separated by an integral number of turns (bottom). The repressor bound at O_{R2} is unable to stimulate transcription efficiently.



in its carboxyl domain also binds noncooperatively to adjacent operator sites, suggesting that the same protein-protein contact mediates the cooperative interaction between repressors bound at adjacent and nonadjacent sites. Assuming repressor to be a rigid structure, it is difficult to imagine how the same surface of the carboxyl domain could be used in both cases (compare Fig. 1 and 6). The results thus suggest that repressor's carboxyl domain is tethered to its amino domain with sufficient flexibility to allow identical interactions for the molecules bound to either adjacent or separated operator sites. We do not know whether the cooperativity mutant is also impaired for dimer formation. This function must be at least partly intact, however, because the mutant repressor confers immunity at a concentration at which the amino domain alone (which dimerizes very weakly) does not.

Why is repressor bound at O_{R2} unable to stimulate transcription efficiently when it interacts with a distally bound repressor? One possible explanation is that the interaction between the two separated repressors constrains repressor bound at O_{R2} , so that the appropriate surface on its amino domain cannot effectively contact RNA polymerase. This explanation would seem to be inconsistent with the evident flexibility of the repressor molecule, but it is possible that a subtle shift in the repressor-operator geometry would be enough to disrupt the contact with bound polymerase. An alternative possibility is that some aspect of the looped structure *per se* interferes with activation. Possibly distortion of the DNA prevents RNA polymerase from interacting properly with the promoter, or perhaps the loop sterically inter-

feres with polymerase. In any case, the following observation implies that the inhibitory effect of the distal repressor depends on the detailed nature of the polymerase-promoter interaction: substitution of a hybrid promoter for P_{RM} partially relieves this inhibitory effect on stimulated transcription (unpublished observation). The hybrid promoter, created by fusing the -35 region of P_{RM} to the -10 region of the *lac UV5* promoter, differs both in sequence from wild-type P_{RM} and also in the spacing between the -35 and the -10 regions.

As mentioned in the introduction, a similar mechanism has already been proposed by Schleif and coworkers^{12,13} to explain the repression of basal transcription of the arabinose operon, but Lee *et al.*^{21,22} have recently described a different process in which this mechanism would not be involved. We imagine, nevertheless, that there could be other cases in which such a mechanism obtains; one, for example, might be repression of the silent mating type locus in yeast, since the same protein has been shown to bind to both the silencer and the activator binding site²³⁻²⁵. The form of negative regulation that we have described involves the interaction between two identical proteins, but perhaps there are other cases in which this form of negative regulation is effected by an interaction between two different proteins. This would appear to be an attractive mechanism for modulating the activity of a positive regulator.

We thank Robin Wharton and Rick Bushman for helpful discussions and for plasmids; Grace Gill, Howard Nash and Ira Herskowitz for comments on the manuscript, and Marie Drotter for purifying the mutant repressor protein.

Received 2 September; accepted 14 October 1988.

1. Ptashne, M. *Nature* **332**, 697-701 (1986).
2. Hochschild, A. & Ptashne, M. *Cell* **44**, 925-933 (1986).
3. Griffith, J., Hochschild, A. & Ptashne, M. *Nature* **322**, 750-752 (1986).
4. Johnson, A. D., Meyer, B. J. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5061-5065 (1979).
5. Irani, M. H., Orosz, L. & Adhya, S. *Cell* **32**, 783-788 (1983).
6. Kuhnke, G. *et al.* *EMBO J.* **5**, 167-173 (1986).
7. Dandanell, G. & Hammer, K. *EMBO J.* **4**, 3333-3338 (1985).
8. Valentin-Hansen, P., Albrechtsen, B. & Love Larsen, J. E. *EMBO J.* **5**, 2015-2021 (1986).
9. Eismann, E., v. Wilcken-Bergmann, B. & Muller-Hill, B. *J. molec. Biol.* **195**, 949-952 (1987).
10. Borowiec, J. A., Zhang, L., Sasse-Dwight, S. & Gralla, J. J. *J. molec. Biol.* **196**, 101-111 (1987).
11. Sasse-Dwight, S. & Gralla, J. D. *J. molec. Biol.* (in the press).
12. Dunn, T. M., Hahn, S., Ogden, S. & Schleif, R. F. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5017-5020 (1984).
13. Martin, K., Huo, L. & Schleif, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3654-3658 (1986).
14. Johnson, A. D. *et al.* *Nature* **294**, 217-223 (1981).
15. Ptashne, M. *A Genetic Switch* (Cell Press, Palo Alto and Blackwell Publishing, 1986).
16. Meyer, B. J., Maurer, R. & Ptashne, M. *J. Molec. Biol.* **139**, 163-194 (1980).
17. Guarente, L., Nye, J. S., Hochschild, A. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2236-2239 (1982).
18. Hochschild, A., Irwin, N. & Ptashne, M. *Cell* **32**, 319-325 (1983).

19. Bushman, F. D. & Ptashne, M. *Cell* **54**, 191-197 (1988).
20. Chadwick, P., Pirotta, V., Steinberg, R., Hopkins, N. & Ptashne, M. *Cold Spring Harbor Symp. quant. Biol.* **35**, 283-294 (1970).
21. Lee, N., Francklyn, C. & Hamilton, R. P. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8814-8818 (1987).
22. Hamilton, E. P. & Lee, N. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1749-1753 (1988).
23. Brand, A. H., Micklem, G. & Nasmyth, K. *Cell* **51**, 709-719 (1987).
24. Shore, D. & Nasmyth, K. *Cell* **51**, 721-732 (1987).
25. Buchman, A. R., Kimmerly, W. J., Rine, J. & Kornberg, R. D. *Molec. cell. Biol.* **8**, 210-225 (1988).
26. Backman, K. & Ptashne, M. *Cell* **13**, 65-71 (1978).
27. Pabo, C. O., Sauer, R. T., Surtevant, J. M. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1608-1612 (1979).
28. Sauer, R. T., Pabo, C. O., Meyer, B. J., Ptashne, M. & Backman, K. C. *Nature* **279**, 396-400 (1979).
29. Pabo, C. O. & Lewis, M. *Nature* **298**, 443-447 (1982).
30. Dente, L., Cesareni, G. & Cortese, R. *Nucleic Acids Res.* **11**, 1645-1655 (1983).
31. Wharton, R. P. & Ptashne, M. *Nature* **326**, 888-891 (1987).
32. Yu, X.-M. & Reznikoff, W. S. *Nucleic Acids Res.* **12**, 5449-5464 (1984).
33. Johnson, A. D., Pabo, C. O. & Sauer, R. T. *Meth. Enzym.* **65**, 839-856 (1980).
34. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-540 (1980).
35. Amann, E., Brosius, J. & Ptashne, M. *Gene* **25**, 167-178 (1983).

Detection of statistical gravitational lensing by foreground mass distributions

Rachel L. Webster*, Paul C. Hewett†, Margaret E. Harding‡ & Gary A. Wegner‡

* Astronomy Department and Scarborough College, University of Toronto, M5S 1A1, Canada

† Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

‡ Dartmouth College, New Hampshire 03755, USA

Physical associations between galaxies and quasars are known to exist at low redshift, but here evidence is presented for the detection of distant quasars associated with foreground galaxies. The quasars which are juxtaposed with faint galaxies are compared with a control sample of quasars selected using identical techniques. We find clear statistical evidence for a greater number of quasar-galaxy juxtapositions than would be expected from random alignments. A possible explanation is gravitational lensing of the background quasars by matter associated with the foreground galaxies. From the enhancement of the quasar surface density in the vicinity of the galaxies we are able to draw some general conclusions about the distribution of the lensing matter, and we show in particular that the mass associated with the foreground galaxies must be substantially greater than is conventionally attributed to a luminous galaxy and its halo.

Statistical lensing is taken here to refer to cases for which significant enhancement of the quasar brightness has occurred, but in which multiple images are either not present or their separations are such that they lie below our resolution limit. Our data cannot distinguish directly between gravitational lensing by discrete compact objects and lensing by an equivalent smooth mass distribution, but we include both in our definition of statistical lensing. We assume that quasar and galaxy redshifts are predominantly cosmological in origin and will not consider theories of the type invoked by Arp².

The data have been compiled from an area of 214 square-degrees using Automated Plate Measuring facility (APM) scans of plates in nine UK Schmidt Telescope fields. The objective-prism spectra of all images brighter than $B_J \approx 18.7$ in this region—irrespective of their morphological appearance—have been surveyed for possible quasar characteristics using a series of simple criteria such as colour relative to common galactic stars, emission or absorption features or strong continuum breaks³. Higher-resolution spectroscopy was then used to select

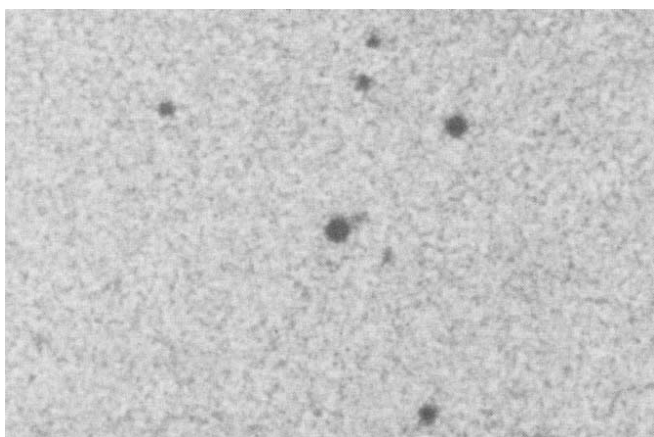


Fig. 1 A typical quasar-galaxy projection reproduced from a B_J direct plate (courtesy of the UKST unit). The quasar redshift is $z = 1.40$. The size of field is approximately 2×1.5 arcmin.

the quasars from among the candidates. A number of quasars show associated structure, such as multiple images, image elongation or regions of low surface brightness ('fuzz')⁴. Using the results of these searches we have formed two subsamples: (1) the control sample of 285 quasars which appear stellar on the UK Schmidt plates, and (2) a sample of 11 quasars which have a faint galaxy within a ~ 6 arcsec radius. These galaxies would be detected as independent images in the APM object catalogue if they were not associated with a quasar. Figure 1 shows one of the quasar-galaxy projections. Redshifts and APM magnitudes are available for all of the quasars. The APM magnitude is directly proportional to the apparent B_J magnitude with a slope of ~ 1 . Both subsamples have been restricted to quasars with redshifts $z > 0.5$ to avoid confusion with galaxy-quasar pairs at the same redshift. It is very unlikely that the galaxies are at the same redshift as the quasars. The surface brightness properties of the galaxies are in fact inconsistent with extended objects at such large redshifts. Furthermore, if the galaxies were at the redshifts of the quasars, the objects would represent an entirely new class of giant superluminous galaxies with luminosities exceeding those of the majority of quasars. More details of the extensive work on galaxies physically associated with low-redshift quasars are given in ref. 1.

We have interpreted our observations in the context of gravitational lensing. The data will be described in detail elsewhere; we are in the process of obtaining a sample ~ 3 times as large, for which we plan to acquire the galactic redshifts soon. Although other examples of quasars with optical structure have been detected⁵, the statistical nature of this survey enables us to evaluate the significance of this phenomenon and draw strong conclusions about the amount of mass in the lensing population.

Typically the galaxies have $B_J \approx 20.5$ – 21.0 , and thus $z_g \approx 0.2$, where we have assumed that the galaxies at the knee of the galaxy luminosity function dominate. To establish the probability of detecting quasar-galaxy projections, we require the optical depth, τ , in visible galaxies on the plate:

$$\tau = \frac{\pi}{360} \left(\frac{R}{6''} \right)^2 \left(\frac{N_g}{10^3} \right)$$

where R is the radius searched around each quasar in arcsec, and N_g is the surface density of galaxies per square-degree. The mean galaxy surface density in the APM object catalogues is $\sim 1,000$ per square-degree, with variations of no more than 10% from field to field. If we search around N_q quasars, we would expect τN_q juxtaposed images, that is,

$$\tau N_q \approx 2.6 \left(\frac{R}{6''} \right)^2 \left(\frac{N_g}{10^3} \right) \left(\frac{N_q}{296} \right)$$

The eleven instances of quasar-galaxy projections that we observe represent a deviation from this prediction at the $>99.99\%$ confidence level, and are thus highly significant. We note that three of the quasars in the control sample were juxtaposed with faint stars. The surface density of stars is very similar to that of galaxies at this limiting magnitude, so that this observation is consistent with our prediction.

Figure 2 shows histograms of the redshift distributions of the two quasar subsamples. A Kolmogorov-Smirnov test confirms that there is no significant difference between the distributions. The redshift distribution of any statistically lensed sample depends on the convolution of the probability of statistical lensing, as a function of quasar redshift for foreground galaxies (out to some limiting redshift), with the intrinsic quasar luminosity function. Using a point mass model for the lens, it can be shown that the variation in the lensing probability for compact objects in galaxies with $z_g < 0.3$ and quasars with $1.0 < z_q < 3.0$ is less than 10% and is therefore effectively constant⁶. Boyle *et al.*⁷ have shown that the quasar luminosity function retains its shape over an extended redshift range. Quasars in our sample

with redshifts $z_q > 0.75$ are drawn from the steep portion of the luminosity function, and thus the shape of the quasar luminosity function is almost constant over the redshift range of interest. A significant difference between the redshift distributions of the two quasar subsamples is therefore not expected. This contrasts with the change in the quasar X-ray luminosity function that has been postulated⁸ to explain the apparent change in the redshift distribution of X-ray-selected quasars juxtaposed with nearby galaxies.

A comparison of the luminosity functions of the subsamples using a Wilcoxon rank-sum test⁹ shows that the two samples are statistically indistinguishable. In other words, the shape of the luminosity functions is the same. The normalization constants of the number-magnitude counts are, however, significantly different. The change in normalization can be estimated by noting that an area defined by a radius of 6 arcsec about each galaxy was searched for quasars. This area is $\sim \pi/360$ of the total surveyed area, and contained eleven quasars, compared with the 285 quasars found in the rest of the surveyed area. Thus the surface density increase is ~ 4.4 , and represents the difference in normalization between the two number-magnitude relations.

To calculate the mean amplification, the slope of the quasar number-magnitude relation is required. Using the comprehensive summary of Boyle *et al.*⁷ and taking into account both the exclusion of quasars with redshifts $z < 0.5$ and our survey's sensitivity to objects with redshifts as high as $z \approx 3.2$, we have taken $d[\log_{10}(N)]/dm \approx -1.0$. Thus the renormalization represents a shift of $\sim 0.6m$ or an average amplification of ~ 1.8 .

It is important to realize that the comparison of the two samples is not sensitive to all forms of lensing, and that both samples may be significantly affected by objects that we cannot detect on the UK Schmidt plates. But we can make an estimate of the effect attributable to statistical lensing by the matter associated with galaxies visible on our plates, as follows. If the galaxies can be modelled as singular isothermal spheres, then $\delta N/N$, the fractional number density enhancement, is proportional to $\theta^{-1}(\sigma_v/c)^2$ for $\delta N/N > 1$, where θ is the radius angle searched and σ_v is the one-dimensional velocity dispersion. The constant of proportionality depends on the slope of the quasar number-magnitude relation, the source size and the mass distribution of compact objects around the galaxy. Using Schneider's terminology and calculations¹⁰,

$$\frac{\delta N}{N} \sim 0.06 f(L) \left(\frac{\sigma_v}{250 \text{ km s}^{-1}} \right)^2 \left(\frac{\theta}{1'} \right)^{-1}$$

If a point source is assumed, the constant of proportionality, $f(L)$ will take its maximum value; more realistic values have been calculated by Schneider¹⁰, and a conservative limit is $f(L) < 1.7$. Then,

$$\sigma_v \geq (450 \text{ km s}^{-1}) \left(\frac{\theta}{6''} \right)^{0.5} \left(\frac{1.7}{f(L)} \right)^{0.5}$$

and we estimate the error on this value to be $\sim 60 \text{ km s}^{-1}$. The velocity dispersion is much larger than that measured for normal galaxies; the mass responsible for the statistical lensing is, therefore, not well modelled by a population of isolated galactic masses. Using a completely different sample, Fugmann¹¹ has also measured an over-density of galaxies near quasars, although he presented no quantitative discussion of the results.

The observed enhancement in the number of quasars about visible galaxies provides a direct measure of the mean surface mass density along the line of sight to the quasars. There are no additional quasar images (at least up to four magnitudes fainter) at distances ≥ 4 arcsec from the galaxies. The absence of such images places significant constraints on the geometry of any multiply imaged quasar configuration. Taking account of these constraints, the observed geometry is inconsistent with the occurrence of macro-lensing by a potential coincident with

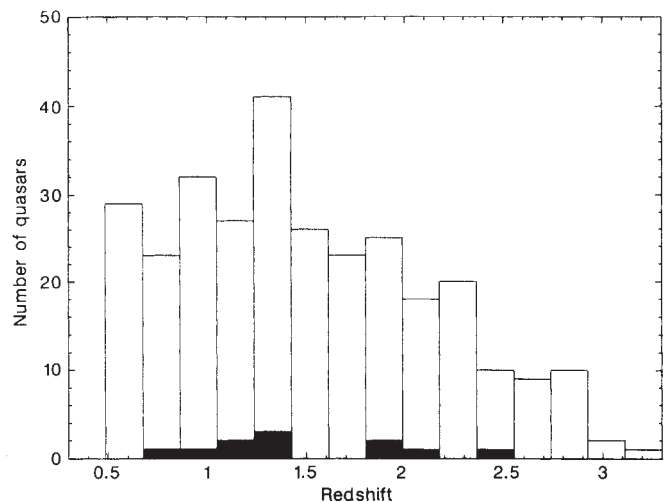


Fig. 2 Redshift histograms of the 285 quasars in the control sample, and the 11 quasars in the statistically lensed sample (solid histogram). Both samples have a redshift cutoff of $z = 0.5$ to avoid confusion with galaxy-quasar pairs at the same redshifts.

the visible galaxy. To reproduce the observations, lensing by any population of objects with a space density significantly lower than that of individual galaxies—such as rich galaxy clusters—must involve very large amplifications. Some instances of multiple imaging (with large image splittings) are expected with high amplifications, and the constraints on the presence of additional quasar images make such a model extremely unlikely.

A possible explanation of the observations invokes mass-clumping on scales of galaxy groups. It is well established that the majority of galaxies exist in groups, and the optical depth, τ , derived above is thus still appropriate. In such a model, the surface density might comprise a component which is part of the visible galaxy, together with a component that is more extended along the line of sight. The extended surface mass density is estimated as $\sim 10^2 M_\odot \text{ pc}^{-2}$, but without information on the spatial extent of the dark matter distribution, the total mass in these systems cannot be determined. This model makes a number of clear predictions—for example, the probability of lensing as a function of redshift. The postulated systems will be effective lenses at redshifts ~ 0.1 – 1.0 , with a maximum at ~ 0.4 . Deep CCD imaging of high-redshift quasars should show an excess of faint galaxies at redshifts larger than those accessible using our plate material.

We thank Scott Tremaine for helpful comments and Israel Kovner for reading the manuscript. Observations described here were obtained with the Multiple Mirror Telescope, a facility operated jointly by the Smithsonian Institution and the University of Arizona. We gratefully acknowledge the cooperation of the UK Schmidt Telescope Unit which provided the plate material, and of the APM group at Cambridge, where the plates were scanned. The APM is funded by the Science and Engineering Research Council as a National Facility. P.C.H. would like to thank the Royal Society for a fellowship. R.L.W. is supported by an NSERC grant. M.E.H. thanks M. F. Ingham for his tolerance.

Received 21 June; accepted 7 October 1988.

1. Yee, H. K. C. & Green, R. *Astrophys. J.* **319**, 28–43 (1987).
2. Arp, H. in *Quasars and Gravitational Lenses* (ed. Swings, J. P.) 307–352 (Université de Liège, 1983).
3. Foltz, C. B. *et al. Astr. J.* **94**, 1423–1460 (1987).
4. Webster, R. L., Hewett, P. C. & Irwin, M. J. *Astr. J.* **95**, 19–25 (1988).
5. Surdej, J. *et al. ESO preprint no. 571* (1988).
6. Turner, E. L., Ostriker, J. P. & Gott, J. R. *Astrophys. J.* **284**, 1–22 (1984).
7. Boyle, B. J., Shanks, T. & Peterson, B. A. *Mon. Not. R. astr. Soc.* (in the press).
8. Stocke, J. T. *et al. Astrophys. J.* **315**, L11–L15 (1987).
9. Lehmann, E. L. *Nonparametrics* (Holden-Day Inc., San Francisco, 1975).
10. Schneider, P. *Astr. Astrophys.* **179**, 80–92 (1987).
11. Fugmann, W. *Astr. Astrophys.* **204**, 73–80 (1988).

Structure of Venus's atmosphere from modelling of night-side infrared spectra

Lucas W. Kamp, Fredric W. Taylor & Simon B. Calcutt

Department of Atmospheric, Oceanic and Planetary Physics,
Oxford University, Clarendon Laboratory, Oxford OX1 3PU, UK

The surface and lower atmosphere of Venus lie below long path-lengths of carbon dioxide and water vapour, and thick cloud layers that were, until recently, thought to be essentially opaque to electromagnetic radiation at wavelengths shorter than a few millimetres. It was unexpected, therefore, when Allen and Crawford¹ announced the detection of measurable quantities of near-infrared radiation leaving the night side. Here we investigate the origin of this radiation by calculating theoretical spectra which we compare with the observations. It is found that the observed radiation can be fully accounted for by thermal emission from the deep atmosphere, and that the intensities suggest surprisingly low abundances for water vapour and carbon monoxide in those layers. These results have implications for probing the atmospheric structure and composition on Venus from the Galileo spacecraft, which is scheduled to make a close encounter with Venus in February 1990.

In view of the high opacity presented by the cloudy atmosphere of Venus to electromagnetic radiation at all but very long wavelengths, the detection of large fluxes leaving the dark side of the planet at certain wavelengths in the near infrared¹ was unexpected. A small amount of sunlight, ~2% of that incident on the planet, is known to diffuse through the clouds to the surface by day; this has been measured by spacecraft landed on Venus by the USSR and the USA. But the thermally emitted radiation of longer wavelengths that leaves the surface, encounters additional absorption by the clouds and also by the vibration-rotation bands of gaseous constituents such as carbon dioxide and water vapour, which are present in large amounts. This 'trapping' of radiant energy from the Sun is responsible for the high surface temperature on Venus, typically ~740 K. The same mechanism, often called the 'greenhouse' effect, has a smaller but still vitally important role in maintaining the more temperate conditions on the Earth.

The night-side emission observed by Allen and Crawford¹, using an instrument on a large ground-based telescope, shows peaks at 1.74 and 2.3 μm , which correspond to gaps between the principal molecular bands of CO_2 . This emission also shows structure across the disk which, despite poor spatial resolution, is reminiscent of that seen in ultraviolet images of Venus on the day side. These facts suggest that the flux originates as thermal emission from the hot surface and lower atmosphere, and is not only transmitted through the entire atmosphere but also preserves a record of the structure of the cloud layers thereby encountered². Therefore, an interpretation of the existing observations can provide new information about a previously almost inaccessible part of Venus's atmosphere. Furthermore, data in this spectral region, including those from forthcoming space missions, may greatly enhance our understanding of the wavelength-dependent transmission of the venusian atmosphere, and hence its composition, cloud structure and greenhouse mechanism. High-spatial-resolution images of the deep cloud structure on Venus may be possible and time-lapsed images could conceivably reveal structure, motions and wave patterns at depths previously hidden from us. In addition, there may be emissions at wavelengths that are not observable from the Earth because of absorptions in our own atmosphere, which would provide further information through the wavelength dependence of the cloud and gas opacities.

To pursue these possibilities we have calculated a set of synthetic spectra for Venus, assuming a model night-side temperature structure taken from the work of Seiff³ and Avduesky

*et al.*⁴. The surface temperature is 740 K. The atmosphere is assumed to be plane-parallel and is divided into 50 layers with logarithmic spacing from heights of 0 to 120 km (outside the clouds), and a cloud zone of 150 layers with linear spacing at heights between 47.5 and 70 km. The pressure structure is obtained by integrating the hydrostatic equilibrium equation, with a surface pressure of 92.0 bar as boundary condition.

We need a spectral range and resolution which covers not only the present ground-based observations, but also those planned from the next spacecraft to visit Venus (see below): the spectral range is 0.7–5.2 μm at a resolution ($\Delta\lambda/\lambda$) of 0.001, which is subsequently degraded to 0.01 using a square filter.

We solve the equation of transfer in the differential-equation form using the method of Feautrier⁵, which allows an exact non-iterative solution that includes both scattering and absorption. We assume a plane-parallel atmosphere and include only coherent scattering and thermal (LTE) emission as sources of radiation, resulting in the following form for this equation⁶:

$$\mu \frac{dI_\nu(\theta, \phi)}{d\tau_\nu} = I_\nu(\theta, \phi) - (1 - \epsilon_\nu) \int_0^{2\pi} \int_0^\pi I_\nu(\theta', \phi') \times P(\theta', \phi', \theta, \phi) \sin \theta' d\theta' d\phi' - \epsilon_\nu B_\nu(T)$$

where θ and ϕ are the polar and azimuthal angles, respectively, $\mu = \cos \theta$, I_ν is the specific intensity at frequency ν , B_ν is the Planck function, P is the scattering phase function, τ_ν is the optical depth and ϵ_ν is the ratio of absorption to total extinction.

The points and weights used for discretization of the polar angle correspond to a 10-point Lobatto quadrature, which gives a point in the nadir direction. The physical characteristics of the cloud particles were taken from Esposito *et al.*⁷, in which scheme only particle modes 2 and 3 (with diameters $>2 \mu\text{m}$) are assumed to contribute at our wavelength range. The total cloud particle extinction coefficient was assumed to be constant over this range, and the absorption coefficient was taken from Tomasko⁸. In the absence of incident sunlight azimuthal symmetry applies. The original formulation by Feautrier relies on the source function being invariant with respect to the sign of the polar angle, and in most of our calculations we use a dipole phase function for cloud particles, which exhibits this symmetry. Tests using a two-component asymmetric⁹ Henyey-Greenstein function showed that the form of the phase function does not significantly affect the shape of the emergent spectrum. A strongly peaked forward or backward scattering function does however, raise or lower (respectively) the emergent intensity, thereby mimicking the effects of varying cloud optical depth (below).

Molecular line absorption was included using data from the AFGL HITRAN line database (Rothman *et al.*¹⁰). The following

Table 1 Mixing ratios of molecules in the 'standard' model

Molecule	Height (km)	Abundance (p.p.m. except where indicated)
CO_2		96.5%
H_2O	<55	100
	55–65	100–2
	>65	2
CO	0–65	0–50
	65–75	50–2
	75–100	2–1,000
	>100	1,000
SO_2	<20	150
	20–56	150–0.05
	>56	0.05
HF		0.005
HCl		0.4
N_2		3.5%

Where a range of values is given, the mixing ratio is varied in a linear fashion over the range of heights listed.

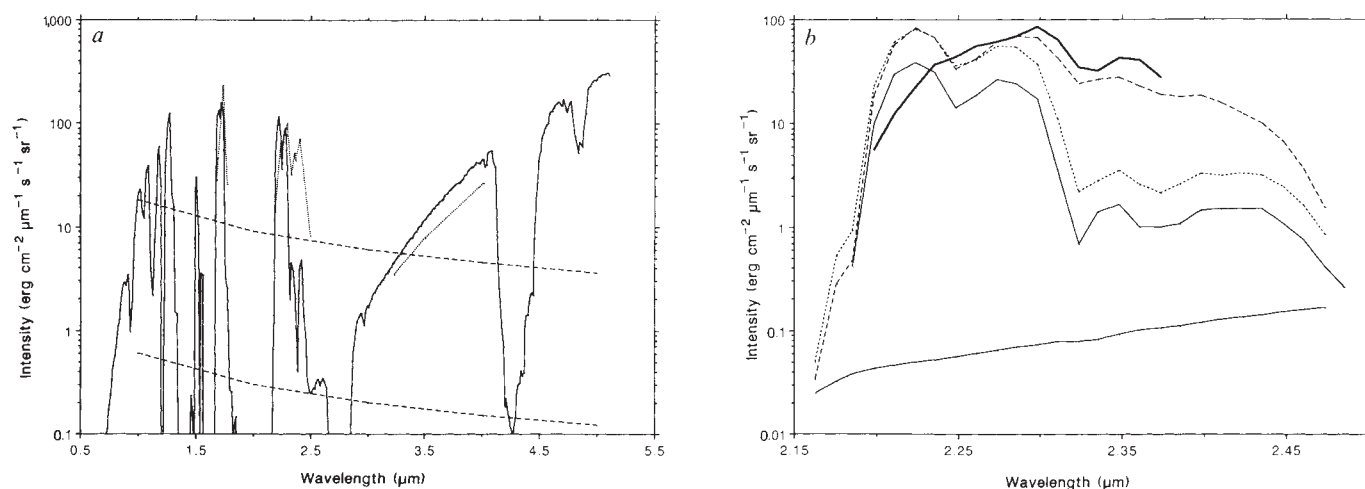


Fig. 1 *a*, Observed and computed spectra for Venus (night-side) in the NIMS range. The solid curve is computed for the standard model; the dotted lines are the ground-based observations of Allen and Crawford¹; the dashed lines represent constant NIMS signal-to-noise ratios of 30 (top) and 1 (bottom). *b*, Spectra in the region of the 2.3- μm peak. The thick solid curve corresponds to the observations of Allen²; the upper thin solid line to the standard model computation; the long-dashed line to a model computation with both H_2O and CO mixing ratios reduced by a factor of ten; the short-dashed line to a model computation with the cloud optical depth reduced by 10%; and the lower thin solid line to a model computation with Lorentzian CO_2 line wings.

species present in the venusian atmosphere have lines in the relevant spectral range: CO_2 , N_2 , H_2O , CO , SO_2 , HF and HCl . Their mixing ratios, taken from refs 11 and 12, are given in Table 1. The line opacities are computed by standard techniques¹³, but special attention needs to be paid to the line profile. The frequency-dependent absorption profile function is usually well represented by a Voigt function, (a convolution of a Doppler (velocity) broadening function with a Lorentzian (pressure) broadening profile). Several studies^{14–16} have shown, however, that in the far wings of infrared CO_2 lines the actual absorption coefficient is considerably below that predicted by the Voigt profile. We expect this to be an important effect for the long path lengths of CO_2 on Venus, especially in the window regions between bands, where the emission peaks of present interest arise and where the gaseous opacity is due mainly to line wings. We therefore include an empirical correction factor, χ , to the Voigt function to allow for the sub-Lorentzian line shape. A reasonable and convenient fit to the data is given by:

$$\chi(\nu, T) = \chi_0 \left[\frac{T_0}{T} \right]^t e^{-C|\nu - \nu_0|}$$

where $C = 0.028$, $\chi_0 = 0.74$ for bands lying above $3,000\text{ cm}^{-1}$, and 3.16 below this, and t varies linearly from 0 at $\Delta\nu = 0$ to 2.0 at $\Delta\nu = 50\text{ cm}^{-1}$, and back to 0 at $\Delta\nu = 170\text{ cm}^{-1}$.

For $\Delta\nu < 1\text{ cm}^{-1}$, χ is set to 1.0 to ensure that this correction factor does not seriously affect the normalization of the profile. No significant effect on the spectrum was found for reasonable variations of these constants, or indeed if the χ -correction was replaced by a simple wing cutoff of 50 cm^{-1} for CO_2 . This lack of sensitivity is reassuring, as this correction factor is highly uncertain. Other possible deviations from standard line-broadening theory, which could result from the high temperatures and pressures of the venusian atmosphere, were assumed negligible in comparison with the uncertainties in cloud radiative properties.

The synthetic spectrum obtained from this (henceforth 'standard') model is shown in Fig. 1*a*, for $\mu = 1$ (nadir viewing). Also shown are the observations of the venusian night side by Allen and Crawford¹, which show good agreement with our results. The spectrum can be interpreted as consisting of essentially two portions. For wavelengths $< 2.5\text{ }\mu\text{m}$, there is thermal emission between the strong CO_2 bands, with peak brightness (black-body) temperatures ranging from 380 K at $2.2\text{ }\mu\text{m}$ to 650 K at $0.89\text{ }\mu\text{m}$. Both peaks observed by Allen and Crawford

are reproduced by us in intensity and approximate shape, although the observed profiles differ in detail from ours. For wavelengths $> 2.5\text{ }\mu\text{m}$, the cloud particle absorption dominates, and the model spectrum approximates a 245-K black body. This is also in quite good agreement with the observations in this range, which correspond to a 240-K black body.

These results support Allen's suggestion² that the emission peaks are due to thermal radiation from the lower atmosphere, but a gap in the cloud cover is not necessary to explain them. Although the clouds are optically thick, at these wavelengths their extinction is primarily through scattering, and the thermal photons from deeper layers can be propagated up through them by this mechanism. The observed spatial variations¹ may arise simply from variations in the cloud optical depths (compare Fig. 1*b*).

A more detailed view of the $2.3\text{ }\mu\text{m}$ peak region is shown in Fig. 1*b*, which also shows observational data from Allen² and several calculations with varying model parameters. The main features of the observed spectrum are reproduced by theory, except for the absence of the predicted peak at $2.23\text{ }\mu\text{m}$; it would be of interest to determine the source of the opacity at that point. The other major discrepancy with respect to the standard model lies in the region longward of $2.32\text{ }\mu\text{m}$, where the computed spectrum is too low. Significantly better agreement is obtained, however, by reducing the mixing ratios of both H_2O and CO by a factor of ten (long-dashed curve in Fig. 1*b*). The short-dashed curve is the result of decreasing the cloud optical depth, which simply shifts the entire spectrum up, as observed by Allen². The latter effect is similar to that obtained by introducing a strongly forward-peaked cloud particle scattering function; therefore, accurate determination of cloud optical depths will require a more precise model of the physical properties of these particles.

The degree to which the agreement with the observations is dependent on the sub-Lorentzian correction function is illustrated by the lower solid curve in Fig. 1*b*, which shows that for purely Lorentzian wings the emission peak vanishes altogether. This phenomenon is due to the large pressure-broadening that one expects for the dense lower atmosphere of Venus. The fact that reducing H_2O abundance improves agreement with observations may again be evidence of sub-Lorentzian behaviour but this is less well understood than is the case for CO_2 .

The radiation in the $2.3\text{-}\mu\text{m}$ peak originates largely from atmospheric layers at heights no greater than 40 km above the surface. If the surface temperature is increased by 40 K (corre-

sponding to a 50% increase in the emitted radiation at that point), the intensity at the top of the atmosphere increases by 5% at most. Therefore the depths for which information can be obtained by this kind of analysis are not likely to extend down to the surface itself. We do predict emission peaks at shorter wavelengths (down to $0.89\text{ }\mu\text{m}$), yielding information about deeper layers, but these peaks have yet to be experimentally confirmed and may be masked by other effects; for example, the peak at $1.27\text{ }\mu\text{m}$ coincides closely with an O_2 airglow line.

There will shortly be an opportunity to observe Venus from the Galileo spacecraft, which is designed to become an orbiter of Jupiter, but is now scheduled also to make a close fly-by of Venus in February 1990. During the Venus encounter, several hours of observations are planned with the Near Infrared Mapping Spectrometer (NIMS), which is capable of generating image data at about 200 wavelengths in the range $0.7\text{--}5.2\text{ }\mu\text{m}$ (ref. 17). The planetary-scale spatial features seen in the Earth-based data, presumably representing the deep cloud structure, points to the desirability of such observations, which combine spatial and spectral resolution. Sensitivity levels for NIMS (corresponding to signal-to-noise ratios of 30 and 1) are shown in Fig. 1a, and indicate that NIMS should be able to observe a number of spectral features on the night-side of Venus which have not yet been seen from Earth. The high-spatial-resolution

coverage during the encounter will include the night side and the interesting dipole and collar regions near the north pole¹⁸, all of which amounts to a unique opportunity to investigate the structure and processes of Venus's atmosphere.

Received 4 August; accepted 17 October 1988.

1. Allen, D. A. & Crawford, J. W. *Nature* **307**, 222-224 (1984).
2. Allen, D. A. *Icarus* **69**, 221-229 (1984).
3. Seiff, A. in *Venus* (eds Hunten, D. M., Colin, L., Donahue, T. M. & Moroz, V. I.) 215-279 (University of Arizona Press, 1983).
4. Avdueskiy, V. S. et al. in *Venus* (eds Hunten, D. M., Colin, L., Donahue, T. M. & Moroz, V. I.) 280-298 (University of Arizona Press, 1983).
5. Feautrier, P., *C. r. Acad. Sci. Paris* **258**, 3189-3191 (1964).
6. Mihalas, D. *Stellar Atmospheres* 2nd edn, 20-36 (Freeman, San Francisco, 1978).
7. Esposito, L. W., Knollenberg, R. G., Marov, M. Y., Toon, O. B. & Turco, R. P. in *Venus* (eds Hunten, D. M., Colin, L., Donahue, T. M. & Moroz, V. I.) 484-564 (University of Arizona Press, 1983).
8. Tomasko, M. G. in *Venus* (eds Hunten, D. M., Colin, L., Donahue, T. M. & Moroz, V. I.) 604-631 (University of Arizona Press, 1983).
9. Milkey, R., Shine, R. & Mihalas, D. *Astrophys. J.* **202**, 250-258 (1975).
10. Rothmann, L. S. et al. *Appl. Opt.* **26**, 4085-4097 (1987).
11. Von Zahn, U., Kumar, S., Niemann, H. & Prinn, R. in *Venus* (eds Hunten, D. M., Colin, L., Donahue, T. M. & Moroz, V. I.) 299-430 (University of Arizona Press, 1983).
12. Krasnopolsky, V. A. & Parshev, V. A. in *Venus* (eds Hunten, D. M., Colin, L., Donahue, T. M. & Moroz, V. I.) 431-458 (University of Arizona Press, 1983).
13. Edwards, D. P. *SPIE Proc.* **298**, 1-23 (1988).
14. Le Doucen, R., Cousin, C., Boulet, C. & Henry, A. *Appl. Opt.* **24**, 897-906 (1985).
15. Winters, B. H., Silverman, S. & Benedict, W. S. *J. quant. Spectrosc. Radiative Transfer* **4**, 527-537 (1964).
16. Burch, D. E., Gryvnak, D. A., Patty, R. R. & Bartky, C. E., *J. Opt. Soc. Am.* **59**, 267-280 (1969).
17. Taylor, F. W. & Calcutt, S. B. *J.Q.S.R.T.* **32**, 463-477 (1984).
18. Taylor, F. W., Kamp, L. W. & Calcutt, S. B. *Space Res.* (in the press).

Light emission from colliding ice particles

W. D. Keith & C. P. R. Saunders

Department of Pure and Applied Physics, UMIST,
Manchester M60 1QD, UK

The electrification of thunderstorms is the result of charge transfer during interactions between cloud and precipitation particles¹⁻⁴. Many laboratory studies with initially uncharged particles⁵⁻¹⁰ have shown that charge is transferred when ice crystals collide with soft hailstones, the charge magnitudes being highly dependent on cloud conditions. Here we describe laboratory experiments which show that the charge transferred during collisions of this type is adequate to account for thunderstorm electrification, but that the net charge transfer is limited, perhaps as a result of local corona breakdown. Light emission from the corona has now been detected. Relatively large charge transfers may occur during contact, but this charge remains essentially localized at the interacting surfaces long enough for corona discharge to cause some charge to return as the particles separate. These results indicate the need for revised charge transfer values in models of thunderstorm electrification.

It is now well established¹⁻⁴ that charges within thunderstorms are carried on the cloud and precipitation particles, and that charge is transferred when ice crystals collide with soft hailstones. The two oppositely charged particles separate from each other and, because of their different size and drag, move apart in the updraught to provide the observed positive charge towards the top of thunderstorms and negative charge lower down. Many theories have been constructed to explain the nature of the charge transfer mechanism. Some of these propose, for example, that it is due to temperature differences, or to potentials caused by freezing; others are inductive theories which rely on the field itself to drive the charge transfer between two colliding ice particles in a regenerative manner. These latter theories have been considered in detail and have been rejected as being incompatible with field and laboratory observations^{9,11-13}. More recently¹⁴ a revised inductive mechanism has been proposed which may be active in all thunderstorms; it cannot, however,

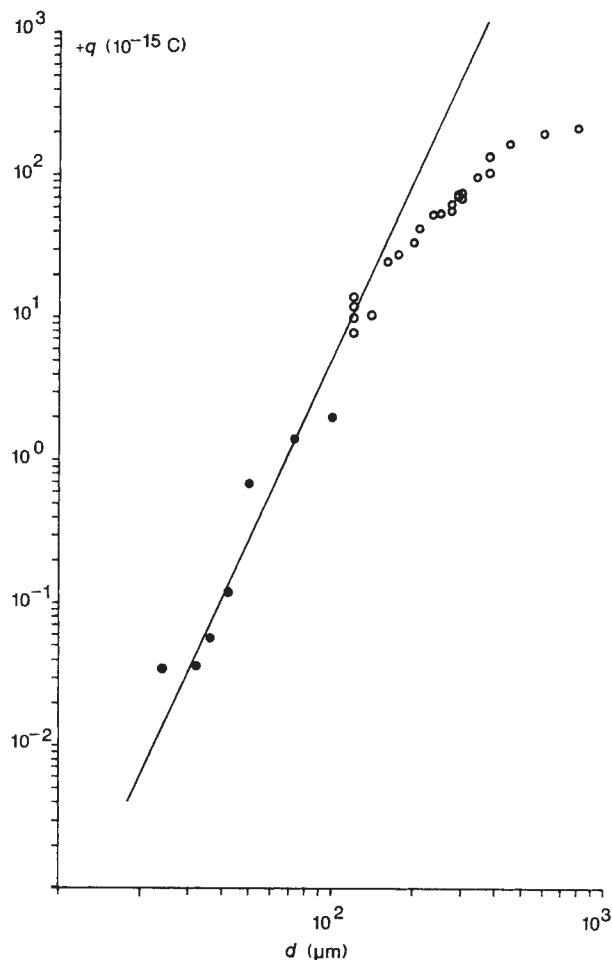


Fig. 1 Positive charge transfer ($+q$) to a riming soft hailstone resulting from interactions with ice crystals of size d . Filled circles are previous results⁹ ($d \leq 100\text{ }\mu\text{m}$) and open circles are the new results discussed here.

account for the observations made in some thunderstorms of large charges on particles¹⁵.

Other laboratory investigations have shown that the surface properties of the two interacting ice particles control the magnitude and the sign of the charge transfer¹⁶⁻¹⁸. It has been proposed¹⁸ that an effective contact potential exists on the surface of the ice particles even though ice is a proton conductor. This interpretation requires, however, that there be an as yet unknown and unmeasured, but specific, contact potential on the surface of ice crystals. Recently, a theory involving different concentrations of charged dislocations on the interacting ice surfaces has been put forward¹⁹ as a possible charge-transfer mechanism.

We have carried out laboratory studies in which charge transfer between initially uncharged ice crystals and soft hailstones has been measured over a range of cloud conditions with no electric field present. In particular, we have extended previous experimental results⁹ by using larger ice crystals, with sizes similar to those found in thunderstorms. Previously, a relationship between charge q and crystal sized of the form $q \propto d^{3.4}$ was found⁹; however, this was derived from crystals of diameter less than $\sim 100 \mu\text{m}$. We find that the $d^{3.4}$ dependence is valid only for diameters up to $\sim 150 \mu\text{m}$, and that this then changes to a $d^{0.3}$ dependence for crystal sizes of at least up to $800 \mu\text{m}$ (see Fig. 1).

The question arises as to why the charge transfer becomes limited for larger crystals. The limit is unlikely to be due to corona discharge over the whole surface of the crystal because, for crystals of this size, the critical charge for corona is $\sim 10^{-12} \text{ C}$, whereas only $\sim 10^{-5} \text{ C}$ is being transferred. We have performed additional experiments at very high interaction speeds, above those relevant to thunderstorms, and find that the charge transferred is orders of magnitude higher than in the experiments at lower velocities. This indicates that at normal collision speeds we have not reached some inherent limit to the amount of charge that ice crystals can hold.

In a series of experiments we used a photodiode to view the zone of interaction between ice crystals and an ice target which simulated a soft hailstone. The crystals were grown from a cloud of supercooled water droplets⁹, and the crystals and droplets were then drawn past the ice target. The apparatus was kept in the dark; when charge transfer was detected, light emission was also found which was two to three orders of magnitude larger than the dark noise. By shielding the photodiode optically but not electrically, the possibility that the output was due to direct electrical pick-up from the corona was excluded. The output obtained when crystals of $100\text{-}\mu\text{m}$ diameter bounced off the target at 8 m s^{-1} corresponds to the emission of 6×10^5 photons per separating crystal. (The concentration of crystals and their collision and separation probabilities were determined separately.) Large resultant charge transfers were found when the cloud contained both supercooled water droplets and ice crystals. In contrast, weak charge transfers were noted when the droplets were absent, so that crystals alone bounced off the target. The light emission followed the same trend, being higher in the presence of liquid water than for crystals alone. Light was emitted for both positive and negative charging of the ice target, the sign of the charge being influenced by the temperature⁹. The charge transfer increased with crystal impact velocity, as did the light output. Light was emitted for all sizes of crystals, but the emission increased with crystal size (Fig. 2). An analysis of the net charge transfer, q , and the associated number of photons emitted, n , revealed a relationship of $n \propto q^{2.5}$ with a correlation coefficient of 0.9.

We suggest that during the glancing contact of an ice crystal with a soft hailstone, so much charge is transferred that when the particles separate, a high electric field is set up which leads to a corona discharge and associated light emission. Despite the low surface conductivity of ice²⁰, there is apparently sufficient contact time for the initial charge transfer to occur but insufficient time for the charges to redistribute over the surfaces

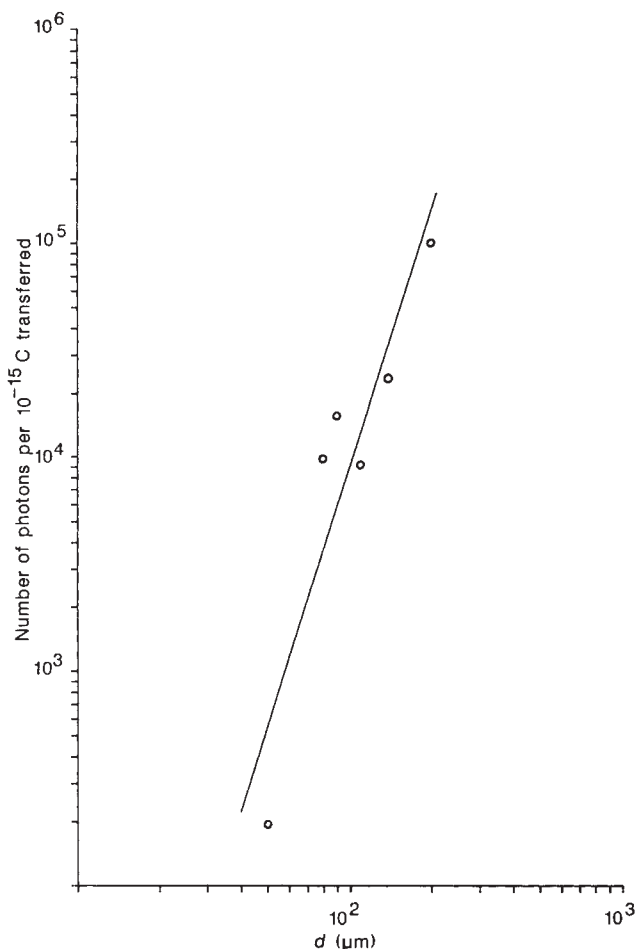


Fig. 2 Average number of photons detected for an individual charge-transfer event per amount of charge transferred, plotted against ice crystal size d .

before the particles separate. Thus every interaction leads to an initial charge transfer and a subsequent reverse charge flow when the particles separate, so that the net transfer appears to be limited.

We can perform an approximate calculation to confirm that the initial charge transfer is adequate to produce a field that is sufficiently high to cause a corona discharge when the particles separate. The initial charge transfer is of course unknown, but it must be larger than the net transfer measured. This transfer can be used to make an estimate of the contact area over which charge is transferred between the particles. Assuming that the crystal and the soft hailstone surface can be considered locally as a pair of flat plates, the Gauss law implies a contact area of $q/\epsilon_0 E$, where q is the net charge transferred and E is the critical electric field across the gap. (During separation, the field remains constant, giving time for the corona discharge to occur.) The limiting electric field at the surface of an ice crystal is of the order of 10^6 V m^{-1} (ref. 21). From Fig. 1, a $100\text{-}\mu\text{m}$ crystal separates a net charge of $10 \times 10^{-15} \text{ C}$, for which the contact area is $\sim 10\%$ of the crystal area. For a $500\text{-}\mu\text{m}$ crystal, the area ratio is 11% , whereas for $50 \mu\text{m}$ it decreases to $\sim 2\%$. Thus, using conservative estimates of the electric charge transferred, the contact areas are considerably smaller than the crystal surface areas. The initial charge transfers require larger contact areas; nevertheless, it is clear that adequate charge is present to lead to corona discharge. The fact that the interacting particles retain a substantial residual charge, rather than losing it all in the discharge process, indicates that there is time for some of the charge originally transferred during the period of contact to leak away from the local area or to become trapped in the

crystal lattice. Obviously, larger initial charge transfers, with larger crystals, lead to more charge being available for the return process. The discharge process limits the net charge transfer by returning some of the charge initially separated before it has time to redistribute in the crystal.

These corona-limited charge transfer processes must be taken into account in any theories of ice/ice-particle charging in thunderstorms as well as in determinations of the net effect throughout a thunderstorm of multiple charge-transfer events. For large crystals the charges actually transferred are less than would be expected from extrapolations of charge-transfer data obtained with small ice crystals. The experimental values of charge transfer obtained as a function of crystal size together with appropriate values of collision and separation probabilities now need to be included in numerical models of thunderstorm electrical development, and the implications of the locally trapped charge on the interacting surfaces should be studied from the theoretical viewpoint of charge transfer in ice.

Received 29 July; accepted 19 October 1988.

1. Krehbiel, P. R., Brook, M. & McCrory, R. A. *J. geophys. Res.* **84**, 2432-2456 (1979).
2. Lhermitte, R. & Krehbiel, P. R. *IEEE Trans. Geosci. Electron.* GE17, 162-171 (1979).
3. Gardiner, B., Lamb, D., Pitter, R., Hallett, J. & Saunders, C. P. R. *J. geophys. Res.* **90**, 6079-6086 (1985).
4. Dye, J. E. *et al.* *J. geophys. Res.* **91**, 1231-1247 (1986).
5. Reynolds, S. E., Brook, M. & Gourley, M. F. *J. Met.* **14**, 426-436 (1957).
6. Church, C. R. thesis, Univ. Durham (1966).
7. Marshall, B. J. P., Latham, J. & Saunders, C. P. R. *Q. Jl R. met. Soc.* **104**, 163-178 (1978).
8. Gaskell, W. & Illingworth, A. J. *Q. Jl R. met. Soc.* **106**, 841-854 (1980).
9. Jayaratne, E. R., Saunders, C. P. R. & Hallett, J. *Q. Jl R. Met. Soc.* **109**, 609-630 (1983).
10. Keith, W. D. thesis, Univ. Manchester (1987).
11. Christian, H. *et al.* *Q. Jl R. met. Soc.* **106**, 159-174 (1980).
12. Gaskell, W. *Q. Jl R. met. Soc.* **107**, 955-966 (1981).
13. Caranti, J. M. & Illingworth, A. J. *J. geophys. Res.* **88**, 8483-8489 (1983).
14. Mason, J. M. *Proc. R. Soc.* **A15**, 303-315 (1988).
15. Saunders, C. P. R. *Weather* **43**, 318-324 (1988).
16. Hallett, J. & Saunders, C. P. R. *J. Atmos. Sci.* **36**, 2230-2235 (1979).
17. Jayaratne, E. R. & Saunders, C. P. R. *J. geophys. Res.* **90**, 13063-13066 (1985).
18. Caranti, J. & Illingworth, A. J. *Nature* **284**, 44-46 (1980).
19. Keith, W. D. & Saunders, C. P. R. *Preprints Atmos. Elec. Conf.*, Uppsala (1988).
20. Caranti, J. M. & Illingworth, A. J. *J. phys. Chem.* **87**, 4078-4083 (1983).
21. Griffiths, R. F. & Latham, J. *Q. J. R. met. Soc.* **100**, 163-180 (1974).

Amorphous oxide precipitates in silicon single crystals

S. Messoloras*, J. R. Schneider†, R. J. Stewart* & W. Zulehner‡

*J. J. Thomson Physical Laboratory, University of Reading, Whiteknights, Reading RG6 2AF, UK

†Hahn-Meitner-Institut, Glienicke Str. 100, D-1000 Berlin 39, FRG

‡Wacker-Chemitronic GmbH, Postfach 1140, D-8263 Burghausen, FRG

Czochralski-grown dislocation-free silicon single crystals¹ are used almost exclusively by the semiconductor industry in the manufacture of contemporary VLSI (very large-scale integration) devices. These crystals contain small concentrations of dissolved oxygen, which diffuses on heating to form precipitates of silica. Because VLSI device performance depends on the size and distribution of such precipitates the characterization of these precipitates is of immense current technological interest. The volume fraction of the precipitates is only of the order of 10 parts per million, which makes their characterization extremely difficult. Here we show that the technique of small-angle neutron scattering can be used to fully characterize, in a single experiment, the precipitates (in terms of orientation, size, number density and volume fraction) in sample volumes of $\sim 10\text{ cm}^3$. This method may be used by the semiconductor industry to characterize the oxide precipitates after device-processing heat treatments in a batch of wafers.

Czochralski-grown dislocation-free silicon crystals¹ contain $5\text{--}18 \times 10^{17}$ interstitial oxygen atoms per cm^3 (giving an atomic oxygen concentration in the range 10-36 parts per million). During crystal growth, oxygen is dissolved from the silica cruc-

ible containing the molten silicon and is incorporated in the material. During the heating cycles necessary for device fabrication the oxygen atoms diffuse and agglomerate to form amorphous SiO_2 precipitates. These precipitates can have disastrous effects if they occur in the active regions of a device. In regions away from the active components, however, precipitates have advantageous effects, because they 'punch out' dislocations which act as 'getters' for the unwanted metallic impurities. In addition, the dissolved oxygen inhibits the dislocation slip that may be produced by differential expansion or contraction during the thermal treatments of the processing stage. Besides the technological significance of oxygen precipitation in silicon, there are still unresolved scientific problems² concerning the oxygen diffusivity, the size and number density of the precipitates and problems associated with the formation of electrically active defects (thermal donors) during heat treatments at around 450°C .

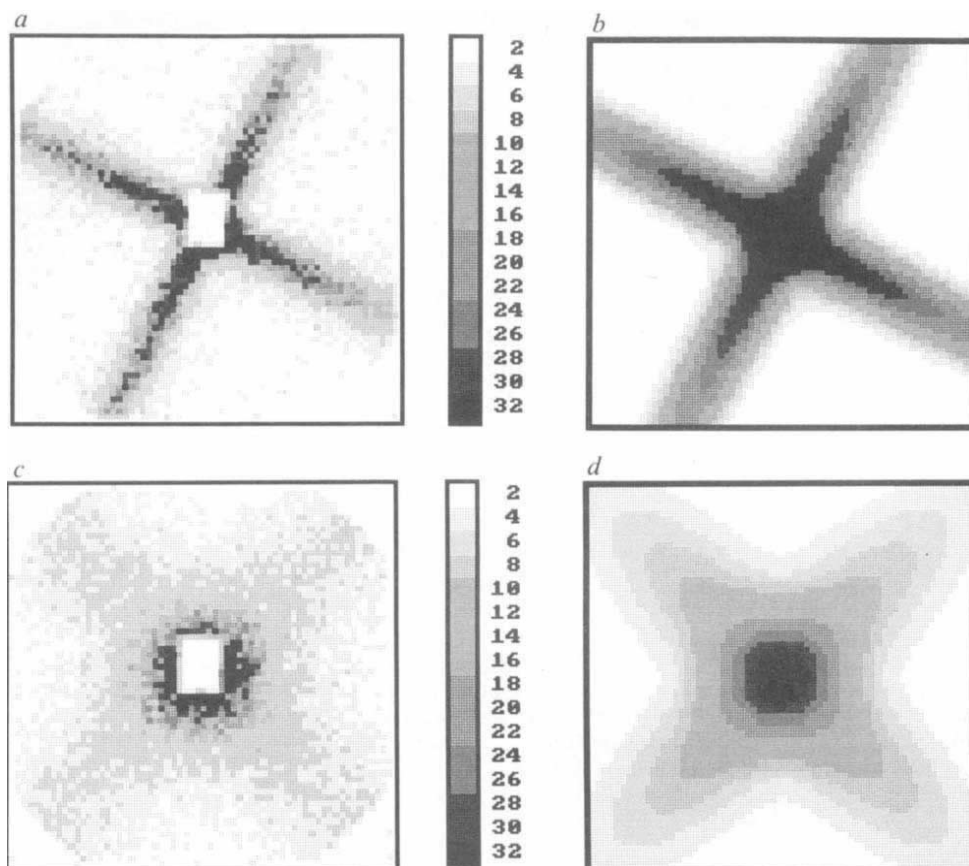
To fully describe the precipitation process of oxygen in silicon, the size, number density, shape and crystallographic orientation of the precipitates must be determined, as well as the concentration of oxygen in the matrix. In the past^{3,4} diverse methods have been used to obtain some of these parameters. Here we demonstrate that they can be all obtained in a single experiment using the technique of small-angle neutron scattering (SANS). Furthermore, this technique has the required sensitivity; it can measure the volume fraction of the oxide precipitates in silicon to within a few parts per million. Another advantage of SANS is that large samples ($\sim 1\text{--}10\text{ cm}^3$) may be used so the parameters obtained are representative of the bulk material.

In this technique neutrons are scattered from inhomogeneities present in a sample; for small-angle scattering sizes in the range $\sim 1\text{--}100\text{ nm}$ can be resolved. In our experiments, the scattering is observed using a square ($64 \times 64\text{-cm}$) two-dimensional position-sensitive detector comprising 3,808 active $1 \times 1\text{-cm}$ cells. The incident neutron beam impinges on the centre of the detector, which is masked by a beam stop. The experimental diffraction patterns (Fig. 1a and c) were obtained using the D11 instrument at the Institut Laue-Langevin, Grenoble, France. The sample-detector distance was 5 m and the incident neutron wavelength was 0.9 nm.

Oriented inhomogeneities give rise to anisotropic scattering, from which the crystallographic orientation of the precipitates can be determined. In Fig. 1a and c, the scattering has a 4-fold symmetry. Because the [100] axis of the sample was parallel to the incident neutron beam we may conclude that the precipitates have cubic symmetry. In Fig. 1a the streaks are extended along the [100] direction, which implies that the short edge of the precipitates is perpendicular to the (100) plane. (The SANS pattern is the Fourier transform of the real object, so the long dimension of the scattering pattern corresponds to the short dimension of the precipitate.) In addition, the lateral extent of the streaks is small, which implies that the other dimension of the precipitate is much larger than that of the [100] direction. In Fig. 1c the streaks are broader, which implies a smaller aspect (length-width) ratio for the precipitates.

In principle, the size of the precipitates could be determined from the Fourier transform of the scattering pattern. This approach, however, presents difficulties, because the data are obtained in only a limited region of reciprocal space. Thus we adopted instead the approach of simulating the scattering. From the symmetry and form of the observed scattering it is reasonable to assume that the precipitates are platelets which lie on {100} planes. This is consistent with the findings of electron microscopy³. A model scattering cross-section is calculated from the Fourier transform of the real-space scattering in this geometry. To fit the model to the experimental data, three parameters can be varied; the large edge of the platelets (L), the small edge (l) and a constant G which is proportional to the volume fraction of the precipitates and to their volume¹. In practice, approximate values for these three variables are found by fitting a model

Fig. 1 Grey-scale coded intensity maps of the neutron scattering from amorphous SiO_2 precipitates grown in silicon single crystals after a heat treatment of 216 h at 750 °C. *a*, *c*, Experiments on samples (1) and (2), described in text. *b*, *d*, Corresponding calculated patterns. The neutron wavelength was 0.9 nm and the sample-to-detector distance was 5 m. In *a* and *c* the centre of the detector is masked by a cadmium beam stop. Intensity scales are shown, in arbitrary units.



cross-section to cuts of the experimental data along the $[100]$ and $[\bar{1}10]$ axes. The calculated scattering cross-sections that exhibited the best fit to the experimental data (Fig. 1*a* and *c*) are shown in Fig. 1*b* and *d*, for platelet precipitates with L parallel to $[110]$ and l parallel to $[001]$. The calculated cross-section is very sensitive to the values of L , l and G , which for a good fit, have to be determined to an accuracy of within 10%.

In these experiments we studied three Czochralski-grown silicon single crystals treated for 216 h at 750 °C. The interstitial oxygen concentrations of these crystals were determined by infrared absorption measurements to be (1) 7.4×10^{17} (Fig. 1*a*), (2) 1×10^{18} (Fig. 1*c*) and (3) 1.8×10^{18} (not shown) oxygen atoms per cm^3 . The carbon concentration was less than 10^{16} atoms per cm^3 . The sizes of the platelets (L , l) for the corresponding fitted models are (1) 50, 5 nm (Fig. 1*b*) and (2) 25, 6.2 nm (Fig. 1*d*). The scattering from sample (3) was isotropic and weaker, indicating that the precipitates were either spheres or cubes. If we assume that they are cubes the analysis of the azimuthally averaged scattering gives a size of $L = l = 11$ nm. Thus, as the concentration of oxygen increases the shape of the silica precipitates changes from a plate-like structure towards a more cubic configuration. From the constant G , the volume fraction of the precipitates may be determined³, and these were found to be (1) 14.8×10^{-6} , (2) 16.5×10^{-6} and (3) 19×10^{-6} . For these determinations it is assumed that the precipitates consist of amorphous silica. The corresponding oxygen concentrations in the precipitates are (1) 5.0×10^{17} , (2) 5.4×10^{17} and (3) 6.3×10^{17} oxygen atoms per cm^3 . Thus we observe, as expected, that a greater number of oxygen atoms are precipitated in the samples that have higher initial oxygen concentration. The number densities of the precipitates were found to be (1) 8.5×10^{11} , (2) 3.0×10^{12} and (3) 1.0×10^{13} precipitates per cm^3 ; again, we observe that a higher initial oxygen concentration results in a higher number density of precipitates.

Because the silicon crystals studied are dislocation-free, the nucleation centres for precipitation will be oxygen agglomerates

formed at early heating times. The mean distance between oxygen atom depends on their number, so the probability of one oxygen atom finding another and forming a dimer after a given diffusion path becomes higher as the oxygen concentration increases. Thus the number density of the nucleation centres, and consequently that of the precipitates, will depend on the initial oxygen concentration. In approximate terms the number density of the dimers initially formed is proportional to c_0^2 , where c_0 is the initial concentration of oxygen in the sample. Thus we expect that the ratio of the precipitate number densities in two samples should be proportional to the square of the ratio of the initial oxygen concentrations in those samples. This is true for samples (2) and (3), and holds within a factor of two for samples (1) and (2). Thus, bearing in mind that the initial stages of nucleation involve more complicated processes than simple dimer formation, this simple model is supported by the results.

We conclude that SANS provides many useful possibilities for the study of oxygen precipitation in silicon. Using this technique, the crystallographic orientation, size and number density and oxygen content of silica precipitates can be determined with considerable precision (we have characterized volume fractions of ~ 15 parts per million). The number density, size and aspect ratio of the platelet precipitates have been shown to depend on the initial oxygen concentration in the crystal. The exact shape of the precipitates cannot be uniquely determined by SANS, however, and thus complementary information from electron microscopy is required^{4,5}.

We wish to thank the Institut Max von Laue–Paul Langevin for the provision of neutron beam facilities.

Received 4 August; accepted 17 October 1988.

- Livingston, F. M. *et al.* *J. Phys. C* **17**, 6253–6276 (1984).
- Messoloras, S., Newman, R. C., Stewart, R. J. & Tucker, J. H. *Semiconductor Sci. Technol.* **2**, 14–19 (1987).
- Binns, M. J. *et al.* *Appl. Phys. Lett.* **42**, 525–527 (1983).
- Bourret, A., Thibault-Desseaux, J. & Seidman, D. N. *J. appl. Phys.* **55**, 825–836 (1984).
- Bergholz, W. *et al.* *Phil. Mag.* (in the press).

Calorimetric evidence for the micro-quasicrystalline structure of 'amorphous' Al/transition metal alloys

L. C. Chen & F. Spaepen

Division of Applied Sciences, Harvard University,
Cambridge, Massachusetts 02138, USA

The idea that a material exhibiting broad diffraction halos does not have a characteristic amorphous structure but is just an assembly of randomly orientated microcrystals is a remarkably resilient one. Invariably it is resurrected when a new type of amorphous material is discovered, only to be put aside in most cases in favour of a truly amorphous structure (one without any lattice periodicity). For fused silica and the amorphous elemental semiconductors, this structure is the continuous random network¹; for the amorphous metals it is a dense random packing of hard spheres (see ref. 2; the strong short-range order in amorphous metals is explicitly taken into account in the 'stereochemically defined' models described in ref. 3). Nonetheless, truly microcrystalline materials do exist, and their diffraction patterns consist of broadened powder peaks⁴. Here we show that these microcrystalline materials can be identified unambiguously by the characteristic isothermal calorimetric signal associated with their transformation to a large-grained material, and we use this test to demonstrate that two sputtered 'amorphous' aluminum/transition metal alloys have a micro-quasicrystalline structure.

Based on electron microscopy and diffraction observations, Bendersky and Ridder⁵ have suggested that rapidly quenched $\text{Al}_{86}\text{Mn}_{14}$ droplets are micro-quasicrystalline, with a quasicrystal size of the order of 1 nm for the highest quench rates (10^6 K s^{-1}). Robertson *et al.*⁶ have shown that the X-ray diffraction pattern of 'amorphous' sputtered $\text{Al}_{72}\text{Mn}_{22}\text{Si}_6$ could be well fitted by broadening the quasicrystalline powder pattern from the same material sputtered at a higher temperature with a combination of size (2.5 nm diameter) and phason strain functions. For the present purposes, quasicrystals are no different from regular crystals, in the sense that the boundary between two quasicrystals of different orientation is as well defined geometrically as a regular grain boundary.

Because the pair distribution function obtained from diffraction experiments on amorphous materials does not, by itself, allow unambiguous identification of the structure, supplementary information from other experiments is necessary. Calorimetry has traditionally been extremely valuable for this purpose. For example, the thermal manifestation of a glass transition is strong evidence for a liquid-like structure. If a glass transition is not observed, however, this does not necessarily mean that the structure is microcrystalline, as the transition can be obscured by crystallization. In that case it is still possible to use calorimetry to identify a microcrystalline structure unambiguously, by observing the isothermal crystallization kinetics.

The 'crystallization', by which we mean the exothermic transformation to a structure with sharp diffraction rings, of truly amorphous materials is qualitatively different from that of microcrystalline ones. In the former case, new crystals nucleate and grow, and the heat given off is the difference in enthalpy between the amorphous and crystalline phases. A microcrystalline assembly, however, does not undergo a phase transformation, but simply coarsens by a process of grain growth, and the heat given off corresponds to the reduction in interfacial enthalpy. The kinetics of these two processes are fundamentally different.

The fraction of the material transformed by an isothermal nucleation-and-growth process can be written as $x = 1 - \exp(-bt^n)$, where b is a constant that depends on the

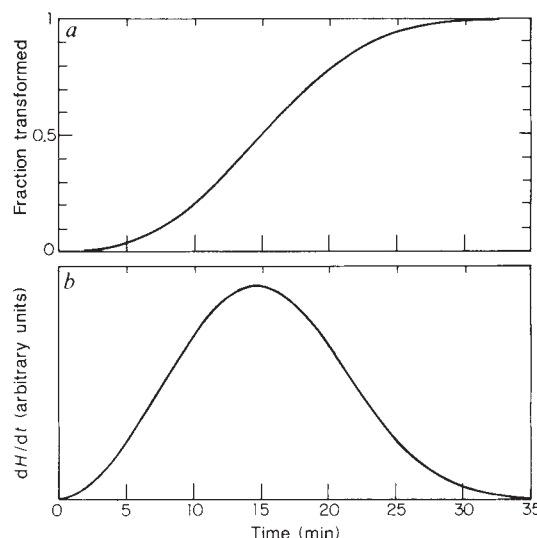


Fig. 1 Simulation of crystallization by isothermal nucleation and growth. *a*, Fraction crystallized. *b*, Corresponding enthalpy release rate. The formalism used is as discussed in the text. The parameters are $n = 2.8$, $K = 3.48 \times 10^{-4} \text{ min}^{-n}$, as determined by Greer⁹ for the crystallization of glassy $\text{Fe}_{80}\text{B}_{20}$.

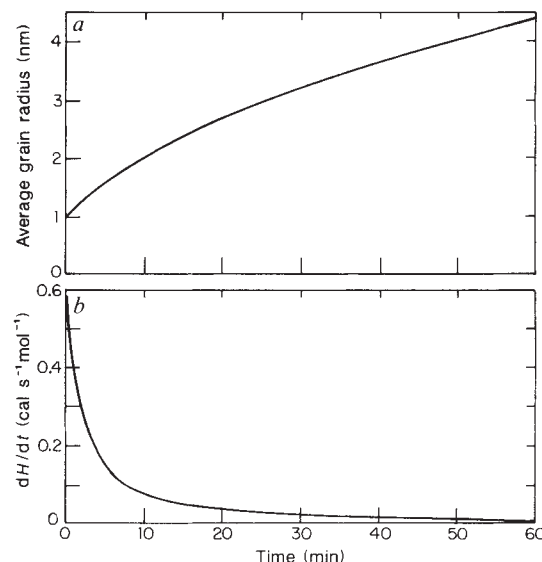


Fig. 2 Simulation of the transformation of a microcrystalline sample by isothermal grain growth. *a*, Evolution of the average grain size. *b*, Corresponding enthalpy release rate. The formalism used is discussed in the text. The parameters are: $m = 1$, $K_0 \lambda^4 \gamma / k = 13 \times 10^3 \text{ m}^2 \text{ K s}^{-1}$, $r(0) = 1 \text{ nm}$, $g\gamma = 0.105 \text{ N m}^{-1}$, and $Q = 241 \text{ kJ mol}^{-1}$. The initial grain radius was taken to be similar to the estimate of Robertson *et al.*⁶. The activation enthalpy, Q , was determined independently from a Kissinger analysis of the peak temperatures at different scan rates^{8,13}. Although it is not obvious that the Kissinger method is applicable to grain growth peaks, analysis and simulation show this to be the case (L.C.C. and F.S., in preparation).

nucleation frequency and the growth velocity, t is the time and n is an exponent, which, depending on the nucleation mechanism, lies between 2 and 4 (ref. 7). Figure 1*a* shows the characteristic sigmoidal shape of the $x-t$ relationship. The heat given off per unit time in such a process is then $-dH/dt = \Delta H(1-x)nbnt^{n-1}$, where ΔH is the total enthalpy of transformation of the sample. It is clear, as illustrated in Fig. 1*b*, that the

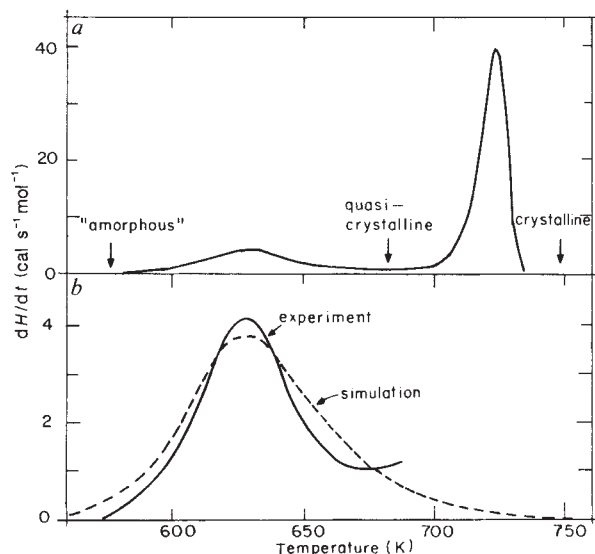


Fig. 3 *a*, Differential scanning calorimeter (DSC) scanning trace from a sputtered $\text{Al}_{82.6}\text{Mn}_{17.4}$ film, heated at 40 K min^{-1} . *b*, Fit of the first peak of *a* with the grain growth formalism discussed in the text. The fit parameters are the same as those of Fig. 2, except that $K_0\lambda^4\gamma/k = 3.3 \times 10^3 \text{ m}^2 \text{ K s}^{-1}$ and $g\gamma = 0.179 \text{ N m}^{-1}$. The experimental curve does not reach zero at higher temperatures because of the onset of the transformation to the stable crystalline phase.

isothermal calorimetric signal for this process is an exothermic peak with a maximum at a non-zero time $t = [(n-1)/bn]^{1/n}$. Qualitatively, in a nucleation-and-growth process the transformation starts at a finite number of points, so that the amount of material transformed at early times is always small. Calorimetric evidence for this type of transformation has been observed and modelled in many systems^{8,9}.

In a grain growth process, the rate of increase of the average grain radius, r , is generally written as $dr/dt = M\gamma/r^m$, where M is a mobility, γ is the interfacial surface tension and m is an exponent, empirically taking a value between 0.5 and 3 (ref. 10). Most theoretical grain growth models¹¹ give $m = 1$. A simple analysis gives the mobility in terms of microscopic quantities as $M = K\lambda^{m+3}/kT$, where K is the net jump frequency across the boundary, λ is the interatomic distance and k is Boltzmann's constant. The jump across the boundary is thermally activated: $K = K_0 \exp(-Q/kT)$. The average grain radius as a function of time can be written as $r^{m+1}(t) = r^{m+1}(0) + (m+1)M\gamma t$, and is shown in Fig. 2a. The corresponding interfacial enthalpy is $H(t) = H(0)[r(0)/r(t)]$, so the isothermal calorimetric signal for the grain growth process is $-dH/dt = H(0)r(0)M\gamma/r^{m+2}$. As illustrated in Fig. 2b, this is a monotonically decreasing signal, which is qualitatively different from the peaked signal observed in nucleation-and-growth.

Our experiments were performed on $\text{Al}_{82.6}\text{Mn}_{17.4}$ and $\text{Al}_{82-83}\text{Fe}_{17-18}$ films, prepared by magnetron and ion beam sputtering, respectively. After removal from the substrate, the films were tested in a Perkin Elmer DSC II differential scanning calorimeter. On heating, two exothermic transformations are observed (Fig. 3a). X-ray and electron microscope observations showed that the structure of the materials in the as-prepared state was 'amorphous' (that is, they exhibited broad diffraction halos and featureless TEM images), and was quasicrystalline after the first transformation and crystalline after the second (L.C.C., F.S., J. L. Robertson, S. C. Moss and K. Hiraga, to be published). The first peak, with a total enthalpy of transformation of $1,050 \text{ J mol}^{-1}$, can be modelled quite well with the grain growth formalism discussed above (Fig. 3b).

Conclusive evidence for grain growth was provided by isothermal measurements, such as that shown in Fig. 4. For these, the

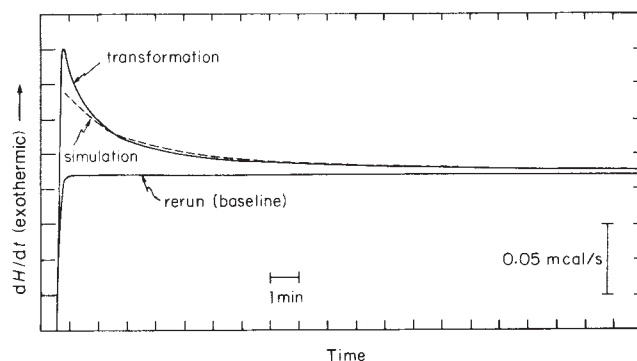


Fig. 4 DSC isothermal trace (solid line) from a sputtered $\text{Al}_{82.6}\text{Mn}_{17.4}$ film, held at 573 K. The dashed line represents the simulation of Fig. 2.

sample is heated at a rate of 20 K min^{-1} to the transformation temperature, and maintained at that temperature. A monotonically decreasing signal (Fig. 2b) characteristic of grain growth is observed, which is qualitatively different from the nucleation-and-growth peak of Fig. 1b. The total enthalpy of the isothermal transformation is 857 J mol^{-1} , which is slightly less than that of the first peak in the scanning experiment, because of growth during the heat-up stage. The possibility that the signal could be the tail end of a hidden transformation peak is therefore excluded. Cooling the fully transformed sample (which is that for which further coarsening is no longer measurable on the timescale of a calorimeter run) to room temperature and re-running the temperature program for the test gave a perfectly horizontal isothermal line (see Fig. 4), ruling out the possibility that the signal is an instrumental artefact. Comparison with the simulation of Fig. 2b shows that the isothermal signal can be modelled quite satisfactorily by the grain growth formalism with parameters similar to those used for modelling the scanning peak (Fig. 3b). That the initial part of the data is more difficult to fit is not surprising, as deviations from the simple macroscopic grain growth formalism are most likely to occur in this regime.

The average interfacial enthalpy can be estimated from the total initial enthalpy $\Delta H(0) = gV\gamma/r(0)$, where g is a geometrical factor that depends on the grain morphology, and V is the volume of the sample. For the Al-Mn sample of Fig. 3a, using $r(0) = 1 \text{ nm}$ as in the fits, this gave $g\gamma = 0.11 \text{ N m}^{-1}$. For typical uniform grain structures $g \approx 1.5$, so the quasicrystalline grain boundary enthalpy seems to be somewhat smaller than typical values for boundaries in crystalline Al (ref. 12).

In conclusion, isothermal calorimeter experiments on 'amorphous' Al-Mn and Al-Fe alloys show a monotonically decreasing evolution of heat that is characteristic of grain growth and is qualitatively different from nucleation-and-growth. It should be kept in mind, however, that the observation of a peak in an isothermal calorimeter experiment does not, by itself, rule out a microcrystalline structure, because abnormal grain growth, in which a few grains grow at the expense of the others, would give a similar signal. Nevertheless, the technique described here constitutes an important addition to the methods available for the identification of the structure of disordered materials.

We thank S. C. Moss and J. L. Robertson for Al-Mn samples and useful discussions, and David Turnbull for a critical reading of the manuscript. This work is supported by the NSF through the Harvard Materials Research Laboratory. L.C.C. acknowledges support by a predoctoral fellowship from IBM Corporation.

Received 3 August; accepted 10 October 1988.

1. Wright, A. C. *J. non-Cryst. Solids* **75**, 15-27 (1985).
2. Cargill III, G. S. *Solid St. Phys.* **30** (eds Ehrenreich, H., Seitz, F. & Turnbull, D.), 227-320 (Academic, New York, 1975).
3. Gaskell, P. H. in *Glassy Metals II; Topics in Applied Physics* **53**, 5-29 (Springer, Berlin, 1983).

4. Wagner, C. N. J., Light, T. B., Halder, N. C. & Lukens, W. E., *J. appl. Phys.* **39**, 3690–3693 (1968).
5. Bendersky, L. A. & Ridder, D. J. *Mater. Res.* **1**, 405–414 (1986).
6. Robertson, J. L., Moss, S. C. & Kreider, K. G. *Phys. Rev. Lett.* **60**, 2062–2065 (1988).
7. Christian, J. W. *The Theory of Transformations in Metals and Alloys* 2nd edn, 525–548 (Pergamon, Oxford, 1975).
8. Henderson, D. W. *J. non-Cryst. Solids* **30**, 301–315 (1979).
9. Greer, A. L. *Acta metall.* **30**, 171–192 (1982).
10. Atkinson, H. V. *Acta metall.* **36**, 469–491 (1988).
11. Thompson, C. V., Frost, H. J. & Spaepen, F. *Acta metall.* **35**, 887–890 (1987).
12. Gleiter, H. & Chalmers, B. *Prog. Mater. Sci.* **16**, 13–42 (1972).
13. Kissinger, H. E. *Anal. Chem.* **29**, 1702–1706 (1957).

Formation of ultrafine-grained magnetite in soils

Barbara A. Maher* & Reginald M. Taylor†

* School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

† CSIRO, Division of Soils, Glen Osmond, S. Australia

The magnetic properties of certain soils indicate the widespread presence of ultrafine-grained magnetite, even where there is no detrital input of magnetite^{1–3}. This suggests that *in situ* formation of magnetite can occur under soil-forming conditions. Schwertmann and Taylor⁴ have stated, however, that magnetite has not been found as a newly formed mineral in soils. Here we report the recovery and identification of pure, ultrafine-grained magnetite from some UK soils that have no apparent external source of magnetite. We consider this magnetite to be of inorganic, *in situ* origin. The identification of magnetite formation during soil development has significance for studies of iron in soils but also has wider environmental implications. Soil-derived magnetite may contribute to the natural remanent magnetism of sediments, and act as a specific indicator of erosional events.

Soil samples were obtained from Exmoor and the south Lake District, UK, from two well-drained, natural brown earth soils developed *in situ*. The soils were selected to exclude any significant detrital input of magnetite; both have weakly magnetic parent materials (Middle Devonian slate and carboniferous limestone, respectively), and no glacial drift is present. Their pH values vary between 5.5–6.0; there was no evidence of charcoal in either profile. Subsamples were taken at 2-cm intervals and were gently crushed and mixed to obtain homogeneity.

The samples were first subjected to a range of mineral (rock) magnetic measurements, to characterize the mineralogy and grain size of the magnetic phases present. Extrinsic magnetic parameters obtained from the air-dried soil samples include low-field magnetic susceptibility (χ), frequency-dependent susceptibility (χ_{FD}), saturation isothermal remanent magnetization (SIRM), anhysteretic susceptibility (χ_{ARM}) and coercivity of remanence ($(B_0)_{CR}$).

In broad mineralogical terms, the magnetic data indicate the predominance of a ferrimagnetic phase. IRM acquisition curves (such as that shown in Fig. 1a) show >90% acquisition of the SIRM at 100 mT, and $(B_0)_{CR}$ values vary only slightly, between 20–28 mT. The grain size of this ferrimagnetic phase can be inferred from the parameters χ_{FD} and χ_{ARM} , which are sensitive to the presence of sub-micron-sized grains, especially those spanning the superparamagnetic (SP)/single-domain (SD) boundary (~ 0.01 – $0.03 \mu\text{m}$)^{6,7}. Figure 1b shows the χ_{FD} (%) and χ_{ARM} (normalized to SIRM) data obtained from the soils; also shown are published data for synthetic magnetites of controlled grain sizes. The χ_{ARM}/SIRM values for the soils vary within the range occupied by those of the synthetic grains of single-domain size or finer. Their χ_{FD} values are closely correlated with the synthetics in the size range $\sim 0.02 \mu\text{m}$. A modified Lowrie–Fuller test was performed, in which we compared the stepwise a.c. demagnetization behaviour of the ARM values (acquired in a d.c. field of 0.04 mT, peak a.c. field of 95 mT) with that of the

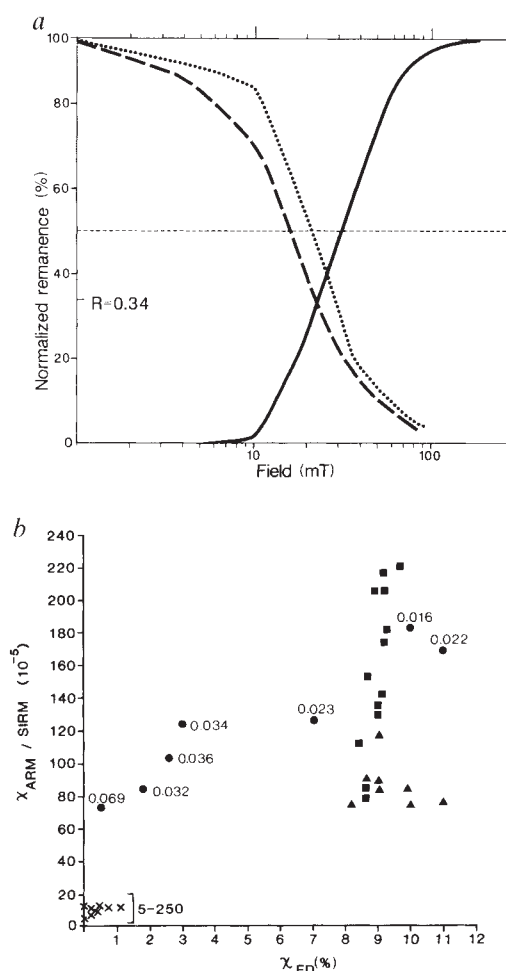


Fig. 1 *a*, Acquisition of IRM (solid line) and a.c. demagnetization of IRM (dashed line) and ARM (dotted line) for the Exmoor brown earth soil, before magnetic extraction. $\text{IRM}_{100 \text{ mT}}/\text{SIRM} = 93\%$, coercivity (at intersection point) = 23 mT, Lowrie–Fuller result ($\text{MDF}_{\text{ARM}} - \text{MDF}_{\text{IRM}} = +3.5 \text{ mT}$, Wohlfarth's 'R' = 0.34 (ref. 20). The IRMs were grown in incremental steps, up to a peak magnetic field of 300 mT, the ARMs in a 0.08-mT direct field and 85-mT-peak alternating field. All remanences were measured using a Digico fluxgate magnetometer with a noise level of $\sim 1 \times 10^{-8} \text{ A m}^2$. *b*, Biparametric plot of the ratio, $\chi_{\text{ARM}}/\text{SIRM}$ versus $\chi_{\text{FD}}(\%)$ for the untreated soils and for synthetic magnetites of documented grain sizes. Here χ_{FD} is the difference between a low-frequency (0.5 kHz) and a high-frequency (5 kHz) measurement of χ expressed as a percentage of the low-frequency measurement. Susceptibility, χ , was measured using a Bartington dual frequency sensor, with a noise level of $\sim 1 \times 10^{-8} \text{ m}^3 \text{ kg}^{-1}$. Numbers by symbols indicate grain sizes in μm . ■, Exmoor brown earth; ▲, Lake District brown earth; ●, Maher (ref. 6); ×, Dankers (ref. 19).

SIRM values. All the soils demonstrate single-domain-type response (that is, Median Destructive Field_{ARM} > MDF_{IRM}). Figure 2 shows the variation of the magnetic parameters with depth in the Exmoor soil. The magnetic content of the soil increases gradually from the base of the soil to the surface, that is, with age of the soil. The SIRM value would indicate a maximum magnetite concentration of $\sim 0.1\%$.

Direct identification of the magnetic mineralogy necessitated extraction and concentration of the magnetic phases (as described below). Curie points were then determined from strong-field thermomagnetic analyses of the magnetic extracts. The thermomagnetic curves (Fig. 3) display Curie temperatures of 580–600 °C. This clearly identifies the soil ferrimagnetic phase as magnetite.

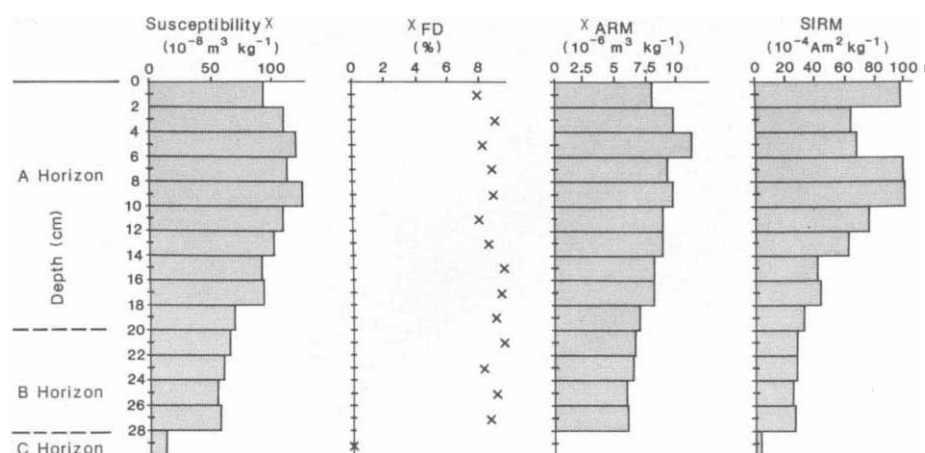


Fig. 2 Variation of magnetic parameters with depth in the Exmoor brown earth soil.

Although the magnetic measurements indicate the presence within the soils of magnetite, apparently consisting of SD- and SP-sized grains, they cannot directly identify an authigenic (*in situ*) origin for this material. Ultrafine-grained magnetite can also be formed both lithogenically⁸ and biogenically^{9,10} within the environment. Other distinguishing properties, notably grain shape characteristics and foreign cation content, must be examined if 'soil' magnetites are to be differentiated from these other known possible sources. This requires direct visual observation and elemental analysis of individual grains of the soil magnetite, extracted from the bulk soil.

We used the extraction technique of Petersen *et al.*⁹. An ultrasonically dispersed soil suspension was circulated around a continuous magnetic extraction system. A rare-earth (iron-neodymium boride) magnet was attached to a soft iron needle, encased in heatshrink plastic tubing, to create a high magnetic field gradient at a point within the system. The aggregates of magnetically extracted particles were collected, re-dispersed and re-concentrated several times to increase purity. We prepared the final extracts for observation under transmission electron microscopy (TEM) by dispersing them in acetone, placing a droplet of the suspension onto a copper-coated grid, and letting the sample evaporate to dryness.

Using a Philips EM400 electron microscope, at magnifications from $\times 170,000$ – $280,000$, we obtained images showing idiomorphic single grains typically with octagonal or hexagonal outlines, which vary continuously in size within the range 0.1 to $<0.01 \mu\text{m}$ (Fig. 4a–d). Elemental spectra were obtained using a Link Systems energy-dispersive X-ray analyser. Iron is often the only detectable component of the crystals (Fig. 4e); trace amounts of zinc and manganese have also been observed.

We consider these geometric microcrystals to be of inorganic authigenic origin, for the following reasons. Their morphology and composition are clearly different from those reported for ultrafine magnetites of lithogenic origin, which occur as elongate inclusions within silicate minerals⁸ and are commonly substituted by titanium. Contrasts with the two types of biogenically produced magnetite¹¹ can also be identified. (Lowenstam¹¹ distinguishes between boundary-organized biomineralization, where magnetite formation occurs actively and intracellularly, and biologically induced mineralization, where export of metabolic products induces extracellular mineral formation.) Petersen *et al.*³ showed recently that boundary-organized magnetosomes, formed by magnetotactic bacteria, constitute a significant natural source of ultrafine magnetite. As with the soil magnetite, this bacterial magnetite is commonly substitution-free, but it frequently occurs in chains, occupies a narrow, biologically constrained, SD-grain-size range, and exhibits distinctive hexagonal or teardrop shapes¹¹. Lovely *et al.*¹³ have reported extracellular production of magnetite by an anaerobic iron-reducing organism, designated GS-15. The magnetite grains

thus produced show several similarities to the soil magnetites; they have a broad size distribution (below $0.05 \mu\text{m}$), are substitution-free, and are not arranged in chains. However, whereas the GS-15 cells seem to grow in intimate contact with the precipitated magnetites, we have observed no bacterial cells associating with the soil magnetites. Finally, as noted by Schwertmann¹⁴, ferrimagnetic phases can be produced in soils through transformation of other soil iron oxides (such as goethite) during burning of the surface soil. Soil goethites and haematites are usually Al-substituted¹⁵, and hence give rise to similarly substituted ferrimagnetic forms. Thus, although formation of ferrimagnets certainly can occur through soil combustion, this mechanism cannot account for the magnetites documented here. Taylor *et al.*⁵ have recently performed a series of laboratory experiments designed to simulate possible pathways of soil magnetite formation. Rapid and easy synthesis of pure, ultrafine-grained magnetites (Fig. 4f) was achieved through controlled oxidation of

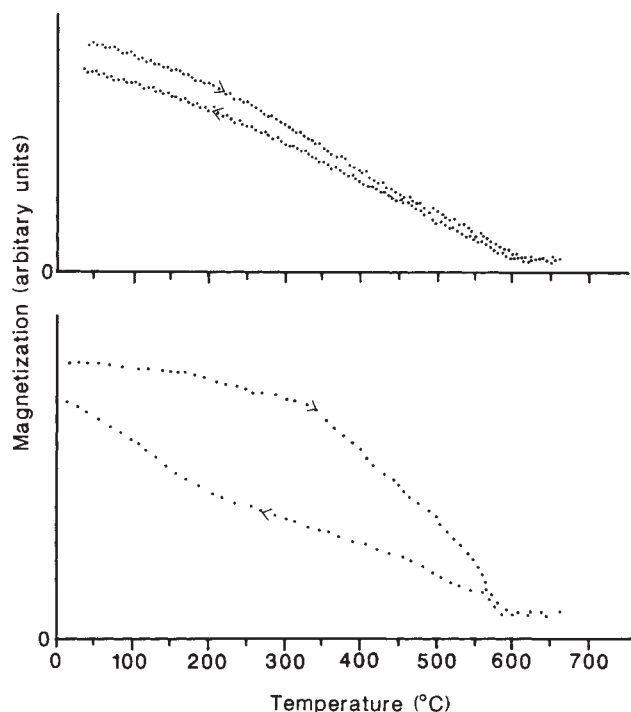


Fig. 3 Curie temperature analyses in air of magnetic minerals extracted by a process of high-gradient magnetic separation from the Exmoor (top) and Lake District (bottom) brown earth soils. Both samples show single-phase Curie points of 580 – 600°C . For both soils, but especially for the Lake District soil, magnetization is lower on the return run, probably indicating conversion of some of the finest-grained magnetite to haematite.

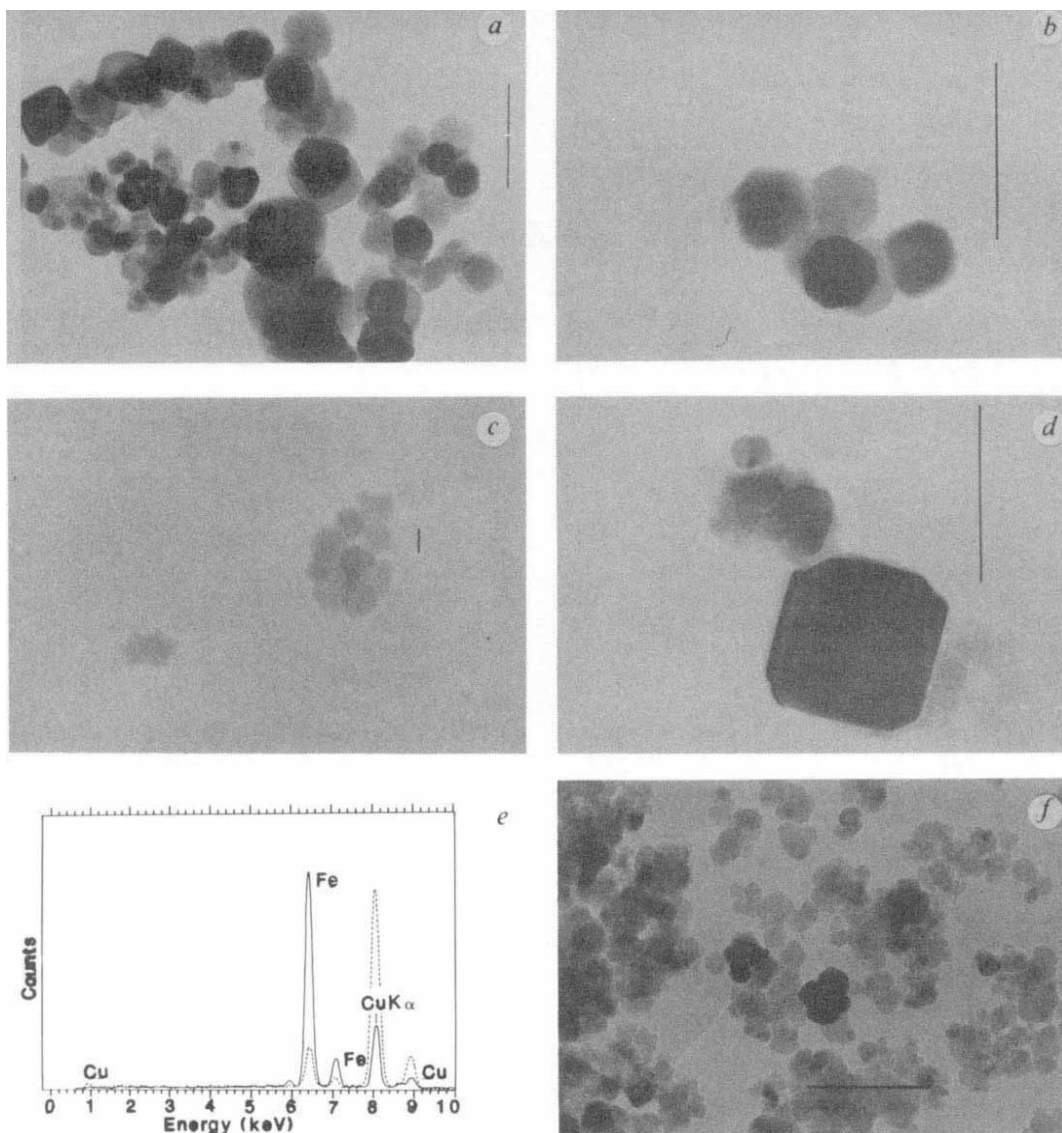


Fig. 4 *a-d*, TEM images of the mineral grains magnetically extracted from the Exmoor and Lake District soils. *a-c* are extracts from Exmoor soil (scale bar in *a* and *b* indicates 0.1 μm , and in *c* it represents 0.01 μm). *d* shows extracts from Lake District soil (scale bar represents 0.1 μm). *e*, Energy-dispersive X-ray analysis of the crystals, obtained in a Philips EM400 with Link Systems analyser (copper peaks are from the support grid). *f*, TEM image of magnetite crystals synthesized at room temperatures and near-neutral pH (scale bar represents 0.1 μm).

Fe^{2+} solutions at room temperature and near-neutral pH (conditions realistic in terms of the soil environment). These synthetic magnetites appear to be close analogues of the soil magnetites, in terms of their magnetic characteristics⁶, their purity, morphology and their grain size.

Single-domain magnetite is resistant to demagnetization and is capable of efficient acquisition of post-depositional magnetic alignment. Even in low concentrations, it may dominate the magnetization of sediments. Soil magnetite, with its mix of SD- and SP-sized grains, may thus contribute significantly to the ultrafine, remanence-carrying magnetic fraction of sediments. It may constitute the source for the magnetites of demonstrably non-biogenic origin observed both by Petersen *et al.*⁹ and by Chang and Kirschvink¹⁶. Furthermore, where it has been identified in sediments, it is often correlated with independent erosional indicators. For example, cores of Recent sediments from the Potomac estuary show major increases in the concentration of ultrafine-grained magnetite during the past 200 years¹⁷. These increases occur synchronously with pollen-analytical evidence of land clearance and agricultural intensification. They are also positively correlated with total sedimentation rates and excess (unsupported) ^{210}Pb activity. The presence of soil magnetite in these, and other, sediments can therefore be used specifically to infer soil erosional inputs¹⁸.

We thank P. Dagley, D. Nichol, T. Cook, G. Bigg and F. Vine for their assistance and helpful discussions. This research was supported by the Leverhulme Trust.

Received 29 July; accepted 14 October 1988.

- Longworth, G. *et al.* *J. Soil Sci.* **30**, 93-110 (1979).
- Maher, B. A. *Phys. Earth planet. Inter.* **42**, 76-92 (1986).
- Mullins, C. E. & Tite, M. S. *J. geophys. Res.* **78**, 804-809 (1973).
- Schwertmann, U. & Taylor, R. M. in *Minerals in Soil Environments* (eds Dixon, J. B. & Weed, S. B.) 145-180 (Soil Sci. Soc. Am., Madison, new Edn in the press).
- Taylor, R. M., Maher, B. A. & Self, P. G. *Clay Miner.* **22**, 411-422 (1987).
- Maher, B. A. *Geophys. J. R. astr. Soc.* **94**, 83-96 (1988).
- Ozdemir, O. & Banerjee, S. K. *Earth planet. Sci. Lett.* **59**, 393-403 (1982).
- Morgan, G. E. & Smith, P. P. K. *Earth planet. Sci. Lett.* **53**, 226-240 (1981).
- Petersen, N., van Dabeneck, T. & Vali, H. *Nature* **320**, 611-615 (1986).
- Lowenstam, H. A. & Kirschvink, J. L. in *Magnetite Biomineralisation and Magnetoreception in Organisms* (eds Kirschvink, J. L., Jones, D. S. & MacFadden, B. J.) (Plenum, New York, 1985).
- Lowenstam, H. A. *Science* **244**, 126-130 (1981).
- Stolz, J. F., Chang, S.-B. R. & Kirschvink, J. L. *Nature* **321**, 849-851 (1986).
- Lovley, D. R., Stolz, J. F., Nord, G. L. & Phillips, E. J. P. *Nature* **330**, 252-254 (1987).
- Schwertmann, U. in *Iron in Soils and Clay Minerals* (eds Stucki, J. W., Goodman, B. A. & Schwertmann, U.) 267-308 (Reidel, Dordrecht, 1988).
- Schwertmann, U. & Fechter, H. *Soil Sci. Soc. Am. J.* **48**, 1462-1463 (1983).
- Chang, S.-B. R. & Kirschvink, J. L. in *Magnetite Biomineralisation and Magnetoreception in Organisms* (eds Kirschvink, J. L., Jones, D. S. & MacFadden, B. J.) 103-151 (Plenum, New York, 1985).
- Oldfield, F., Maher, B. A. & Appleby, P. G. *J. Quat. Res.* (submitted).
- Thompson, R. & Oldfield, F. *Environmental Magnetism* (Allen & Unwin, London, 1986).
- Dankers, P. H. thesis, Univ. Utrecht (1978).
- Cisowski, S. *Phys. Earth planet. Inter.* **26**, 56-62 (1981).

Geochemical evidence for plume–mantle interactions beneath Kerguelen and Heard Islands, Indian Ocean

M. Storey*, A. D. Saunders*, J. Tarney*, P. Leat†, M. F. Thirlwall‡, R. N. Thompson†, M. A. Menzies‡ & G. F. Marriner‡

* Department of Geology, University of Leicester, University Road, Leicester LE1 7RH, UK

† Department of Geological Sciences, University of Durham Science Laboratories, South Road, Durham, DH1 3LE, UK

‡ Department of Geology, Royal Holloway and Bedford New College, Egham, Surrey TW20 0EX, UK

Within-plate ocean island basalts probably originate by the partial melting of upwelling plumes coming from deep within the Earth's mantle¹. Part of the observed heterogeneity in Sr, Nd and Pb isotopic ratios in oceanic basalts may result from the interaction of plumes with the lithosphere^{2–4} and asthenosphere^{2,5}. Kerguelen and Heard Islands, in the southern Indian Ocean, are the latest products of a large and long-lived plume system⁶, comparable to the Hawaiian–Emperor chain in the Pacific. Kerguelen Island began forming ~40 Myr ago on the Antarctic plate, near to the embryonic Southeast Indian Ridge (SEIR), and represents the continuation of the hotspot activity preserved on the Indian plate as the Ninetyeast Ridge⁷ (Fig. 1). With continuing sea-floor spreading, younger volcanism on Kerguelen has occurred further away from the influence of the SEIR. Here we show that the isotopic compositions of Kerguelen basalts vary in sympathy with this changing tectonic environment, and suggest not only that depleted asthenosphere, but also that Cretaceous oceanic plateau lithosphere, may have contributed significantly to the earlier magmatic history of the island.

Kerguelen and Heard Islands are built on the northern half of the older Kerguelen Plateau, above a geoid high which is indicative of upwelling mantle material⁸. Although there have been proposals that the plateau represents a continental remnant of the breakup of Gondwana⁹, recent drilling by the Ocean Drilling Program has confirmed its basaltic nature¹⁰. The Kerguelen Plateau basement appears to be of lower Cretaceous age^{11,12}. Apparently built upon this older basement, the subaerial part of Kerguelen Island largely comprises four massive shield volcanoes consisting predominantly of tholeiitic to transitional basalts¹³. The shield basalts represent the early constructional phase of the island; radiometric dates on associated basic intrusives suggest ages of at least 40 Myr (ref. 14). Younger volcanism comprises alkali basalts and their differentiates, which include eruptions from more localized stratovolcanoes, cinder cones and domes. The youngest of these volcanoes may have been active until late Quaternary times¹³. Heard Island is constructed on middle Eocene to mid-Oligocene pelagic sediments¹⁵. Volcanism has occurred from the late Miocene to the present and has erupted rocks ranging in composition from alkali basalt to trachyte¹⁶. Eroded Quaternary phonolites make up the small McDonald Islands, which lie about 70 km west of Heard Island¹⁶.

The data presented here are based on rocks from Kerguelen and Heard Island collected by HMS *Challenger* in 1874 and by an earlier Antarctic expedition between 1839–43 led by Sir James Clark Ross¹⁷. The majority of samples come from the Loranchet Peninsula (Fig. 1) in the north of the island, regarded as one of the oldest shield volcanoes on Kerguelen¹³, but the younger alkaline flows on Kerguelen are also represented, as are basalts from Heard Island. The samples thus encompass a large period of the eruptive history of these islands. From our analyses (Table 1) and from the trace element and isotopic data presented in Fig. 2, it is evident that there are three compositionally distinct

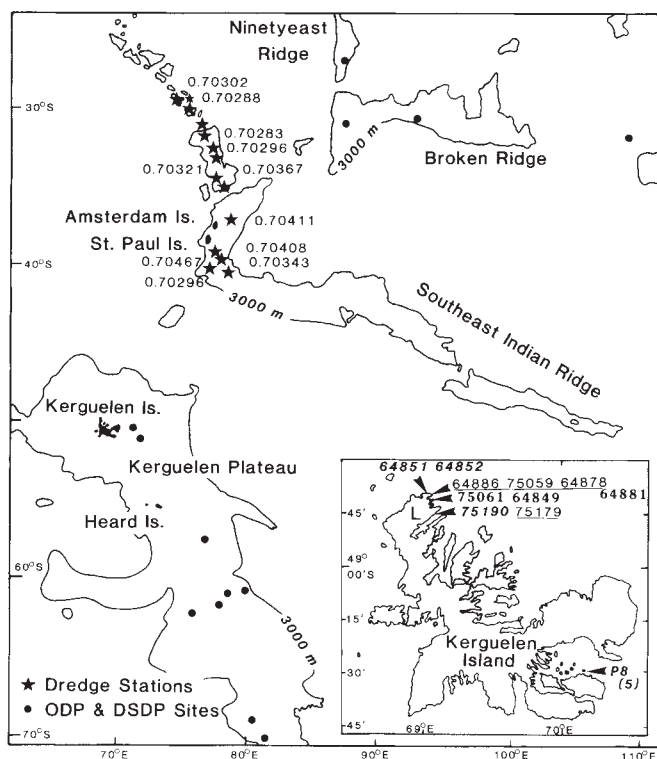


Fig. 1 The location of Kerguelen and Heard Islands with respect to the Kerguelen Plateau, Southeast Indian Ridge, Broken Ridge, Ninetyeast Ridge and St Paul and Amsterdam Islands. The Kerguelen Plateau and Broken Ridge were originally contiguous, but were separated by sea-floor spreading with the formation of the SEIR about 40–45 Myr ago²⁷. Numbers indicate the $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of dredged basalts, which increase towards the SEIR ridge segment containing St Paul and Amsterdam Islands. Isotopic data are from refs 19–21. Inset shows a map of Kerguelen Island and basalt location and type: tholeiites are denoted by bold type, high Na/K alkali basalts by underlined type and low Na/K alkali basalts by italic bold type. Loranchet Peninsula (L).

basalt types amongst the Kerguelen samples (all of which occur in the Loranchet Peninsula). They have the following compositional characteristics:

(Type 1) Tholeiites. These have low incompatible-element abundances, low Zr/Nb (6.9–9.1) and $^{143}\text{Nd}/^{144}\text{Nd}$ (0.512665–0.512778) ratios and high $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (0.70476–0.70499). Kerguelen tholeiitic basalts with lower $^{87}\text{Sr}/^{86}\text{Sr}$ and higher $^{143}\text{Nd}/^{144}\text{Nd}$ ratios than the Loranchet samples occur on Foch Island¹⁸ (Fig. 2), a prolongation of the Cook shield volcano¹³.

(Type 2) High Na/K alkali basalts. These have higher incompatible-element abundances, higher Zr/Nb (11.2–12.5) and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios (0.512790–0.512828) and lower $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (0.70427–0.70463). Relative to Loranchet tholeiites they show a progressive depletion in successively more incompatible elements (Fig. 2a). Their Pb isotopic ratios are close to normal SEIR mid-ocean-ridge basalt (MORB)^{19–21}. There is a continuous range in composition between these rocks and the tholeiites, and as the content of Sr decreases, the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio increases (inset to Fig. 2b).

(Type 3) Low Na/K alkali basalts. These have high incompatible-element abundances but similar highly-incompatible-element ratios to the tholeiites (Fig. 2a). $^{87}\text{Sr}/^{86}\text{Sr}$ ratios are high (0.70541–0.70598) with low Zr/Nb (5.0–5.5) and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios (0.512498–0.512541). A plot of $^{208}\text{Pb}/^{204}\text{Pb}$ against $^{206}\text{Pb}/^{204}\text{Pb}$ for these rocks lies well above that defined by MORB and ocean island basalts from the Northern Hemisphere (Fig. 2c). Most of these appear to post-date the tholeiites and are very similar in composition to Heard Island basalts (Table 1)

Table 1 Major element, trace element and isotope data for Kerguelen and Heard Island basalts

Sample	Loranchet tholeiites				Kerguelen Island				Loranchet alkali basalts (high Na/K)			
	BM 64947	BM 64849	BM 75061	BM 64881	BM 64878	BM 75059	BM 64886	BM 75179	BM 64878	BM 75059	BM 64886	BM 75179
SiO ₂	47.6	47.0	47.4	48.6	46.3	46.0	45.2	45.4	46.3	46.0	45.2	45.4
TiO ₂	1.35	2.15	1.45	1.57	3.29	2.44	2.50	2.58	3.29	2.44	2.50	2.58
Al ₂ O ₃	15.4	14.3	16.3	14.8	14.6	14.3	14.0	14.6	14.6	14.3	14.0	14.6
Fe ₂ O ₃	12.0	13.7	11.4	12.8	14.7	13.1	13.0	13.4	14.7	13.1	13.0	13.4
MnO	0.18	0.18	0.17	0.17	0.20	0.18	0.16	0.17	0.20	0.18	0.16	0.17
MgO	12.3	10.1	7.6	10.7	6.7	11.6	12.0	10.4	6.7	11.6	12.0	10.4
CaO	9.6	11.0	11.7	9.9	10.0	8.7	7.3	7.4	10.0	8.7	7.3	7.4
Na ₂ O	1.9	2.2	2.6	2.0	3.0	3.2	3.3	3.9	3.0	3.2	3.3	3.9
K ₂ O	0.55	0.33	0.26	0.28	0.77	0.98	1.08	1.19	0.77	0.98	1.08	1.19
P ₂ O ₅	0.16	0.19	0.18	0.14	0.45	0.37	0.61	0.70	0.45	0.37	0.61	0.70
Trace elements (p.p.m.)												
XRF												
Ni	262	242	280	286	86	293	364	282	86	293	364	282
Cr	801	426	681	528	143	484	522	352	143	484	522	352
Nb	12	15	15	7	19	18	23	24	19	18	23	24
Zr	83	118	101	64	237	202	257	272	237	202	257	272
Y	18	25	22	18	36	29	26	28	36	29	26	28
Sr	213	343	271	385	513	768	975	1044	513	768	975	1044
Rb	14	8	4	3	14	15	15	19	14	15	15	19
Ba	187	96	98	69	208	248	415	317	208	248	415	317
Zr/Nb	6.9	7.9	6.7	9.1	12.5	11.2	11.2	11.3	12.5	11.2	11.2	11.3
INAA												
Sc	35.1		36.9	29.3	33.0	23.6	16.8		33.0	23.6	16.8	
Th	1.6	1.0	1.7	0.5	1.1	1.4	2.2		1.1	1.4	2.2	
Hf	2.3	3.0	2.7	1.9	5.2	4.6	5.5		5.2	4.6	5.5	
Ta	0.89	1.01	0.88	0.44	1.10	1.21	1.45		1.10	1.21	1.45	
La	13.2	10.8	17.6	7.4	21	21	31		21	21	31	
Ce	29	24	31	19.4	56	52	79		56	52	79	
Nd	13.3	16.5	15.3	12.6	34	31	51		34	31	51	
Sm	3.0	4.3	3.4	3.3	7.9	6.7	8.7		7.9	6.7	8.7	
Eu	1.03	1.35	1.14	1.13	2.48	2.10	2.63		2.48	2.10	2.63	
Tb	0.52	0.76	0.56	0.56	1.41	1.07	1.02		1.41	1.07	1.02	
Yb	1.66	1.99	1.92	1.66	3.22	2.35	1.98		3.22	2.35	1.98	
Lu	0.24		0.27	0.21	0.45	0.31	0.28		0.45	0.31	0.28	
ID												
U	0.238				0.283	0.316	1.103		0.283	0.316	1.103	
Isotope ratios												
⁸⁷ Sr/ ⁸⁶ Sr	0.70499(1)	0.70494(1)	0.70492(1)	0.70475(1)	0.70459(1)	0.70440(1)		0.70427(1)	0.70459(1)	0.70440(1)		0.70427(1)
⁸⁷ Sr/ ⁸⁶ Sr(a.l.)	0.70499(2)				0.70463(1)				0.70463(1)			
¹⁴³ Nd/ ¹⁴⁴ Nd	0.512672(12)		0.512665(5)	0.512778(10)	0.512790(10)	0.512827(16)	0.512828(8)	0.512824(5)	0.512790(10)	0.512827(16)	0.512828(8)	0.512824(5)
²⁰⁶ Pb/ ²⁰⁴ Pb					18.112	18.313	18.088		18.112	18.313	18.088	
²⁰⁷ Pb/ ²⁰⁴ Pb					15.481	15.592	15.504		15.481	15.592	15.504	
²⁰⁸ Pb/ ²⁰⁴ Pb					38.290	38.477	38.288		38.290	38.477	38.288	
Kerguelen alkali basalts (low Na/K)												
Heard Island alkali basalts												
Sample	BM 1967, P8(5)	BM 75190	BM 64851	BM 64852	BM 1967, P9	BM 64986	BM 64992	BM 64990	BM 1967, P9	BM 64986	BM 64992	BM 64990
SiO ₂	44.9	46.2	46.6	47.5	42.0	45.3	46.5	45.1	42.0	45.3	46.5	45.1
TiO ₂	3.49	1.93	2.08	2.2	4.97	4.46	4.21	3.07	4.97	4.46	4.21	3.07
Al ₂ O ₃	11.7	13.1	13.9	14.3	8.2	11.5	12.9	10.5	8.2	11.5	12.9	10.5
Fe ₂ O ₃	12.7	12.8	12.5	13.1	14.9	13.3	12.9	12.8	14.9	13.3	12.9	12.8
MnO	0.16	0.18	0.18	0.17	0.18	0.17	0.16	0.16	0.18	0.17	0.16	0.16
MgO	12.9	13.6	12.8	10.4	15.4	11.2	8.9	13.9	15.4	11.2	8.9	13.9
CaO	8.6	9.8	9.1	9.4	9.8	9.4	9.0	9.1	9.8	9.4	9.0	9.1
Na ₂ O	2.7	1.8	1.4	1.9	1.8	2.2	2.8	2.2	1.8	2.2	2.8	2.2
K ₂ O	2.31	1.39	1.31	1.51	1.72	2.40	2.36	1.88	1.72	2.40	2.36	1.88
P ₂ O ₅	0.66	0.37	0.39	0.41	0.60	0.73	0.75	0.50	0.60	0.73	0.75	0.50
Trace elements (p.p.m.)												
XRF												
Ni	424	363	281	250	487	299	214	441	487	299	214	441
Cr	516	615	533	483	632	385	312	565	632	385	312	565
Nb	55	34	31	33	54	61	56	47	54	61	56	47
Zr	296	169	159	182	270	337	361	282	270	337	361	282
Y	27	20	20	21	25	27	30	26	25	27	30	26
Sr	706	501	608	449	635	845	839	622	635	845	839	622
Rb	62	48	30	38	34	57	45	41	34	57	45	41
Ba	820	545	373	488	436	764	658	614	436	764	658	614
Zr/Nb	5.4	5.0	5.1	5.5	5.0	5.5	6.5	6.0	5.0	5.5	6.5	6.0
INAA												
Sc	23.0	27.7		29.9	27	23	22.7	24.1	27	23	22.7	24.1
Th	6.1	4.2		4.3	3.4	5.6	5.3	5.5	3.4	5.6	5.3	5.5
Hf	7.1	3.9		4.6	6.5	8.0	8.6	3.03	6.5	8.0	8.6	3.03
Ta	3.6	1.99		2.29	3.6	4.04	3.08	6.6	3.6	4.04	3.08	6.6
La	56	40		40	35	49	54	41	35	49	54	41
Ce	123	79		80	85	105	109	87	85	105	109	87
Nd	54	32		34	45	54	58	43	45	54	58	43
Sm	9.1	5.5		6.1	8.9	10.7	11.9	7.6	8.9	10.7	11.9	7.6
Eu	2.52	1.49		1.72	2.59	3.09	3.20	2.29	2.59	3.09	3.20	2.29
Tb	1.15	0.83		0.85	1.13	1.10	1.20	1.04	1.13	1.10	1.20	1.04
Yb	1.44	1.67		1.96	1.41	1.64	1.87	1.71	1.41	1.64	1.87	1.71
Lu	0.17	0.24		0.27	0.15	0.24	0.27	0.17	0.15	0.24	0.27	0.17
ID												
U		0.693			0.816	1.195	1.030		0.816	1.195	1.030	
Isotope ratios												
⁸⁷ Sr/ ⁸⁶ Sr	0.70541(2)	0.70567(2)	0.70567(2)	0.70598(1)	0.70541(1)	0.70555(1)	0.70575(1)	0.70586(1)	0.70541(1)	0.70555(1)	0.70575(1)	0.70586(1)
⁸⁷ Sr/ ⁸⁶ Sr(a.l.)	0.70543(1)	0.70568(1)	0.70570(1)									
¹⁴³ Nd/ ¹⁴⁴ Nd	0.512541(8)	0.512498(7)		0.512507(13)	0.512597(8)	0.512582(5)			0.512597(8)	0.512582(5)		
²⁰⁶ Pb/ ²⁰⁴ Pb	18.086	18.060				18.009	17.993			18.009	17.993	
²⁰⁷ Pb/ ²⁰⁴ Pb	15.532	15.537				15.547	15.559			15.547	15.559	
²⁰⁸ Pb/ ²⁰⁴ Pb	38.710	38.884				38.461	38.511			38.461	38.511	

For isotope ratio data, figures in brackets indicate $\pm$ errors in the final digit or digits. XRF—X-ray fluorescence; INAA—instrumental neutron activation analysis; ID—isotope dilution. Errors for Sr and Nd isotope data are 2σ internal errors; SRM987 gave 0.710242(20) (2σ , $N=73$) and a laboratory Nd standard gave 0.511419(11) (2σ , $N=45$), equivalent to La Jolla of 0.511858(7) during the period of analysis. Isotopic ratios given are measured values where $^{87}\text{Sr}/^{86}\text{Sr}$ was normalized to $^{86}\text{Sr}/^{88}\text{Sr}=0.1194$ and $^{143}\text{Nd}/^{144}\text{Nd}$ to $^{146}\text{Nd}/^{144}\text{Nd}=0.7219$. a.l. = acid leached (6M HCl) in teflon capsule for 2–3 hours at 185 °C. Within-run errors on the Pb isotope data are better than 0.005% per a.m.u. and are normalized to SRM981, which is reproducible to better than 0.04% per a.m.u.

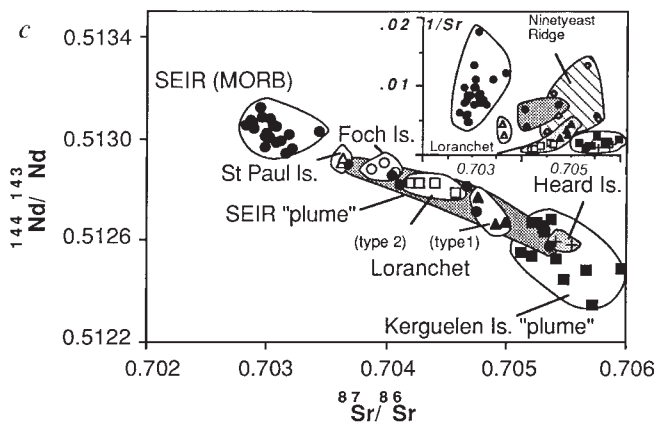
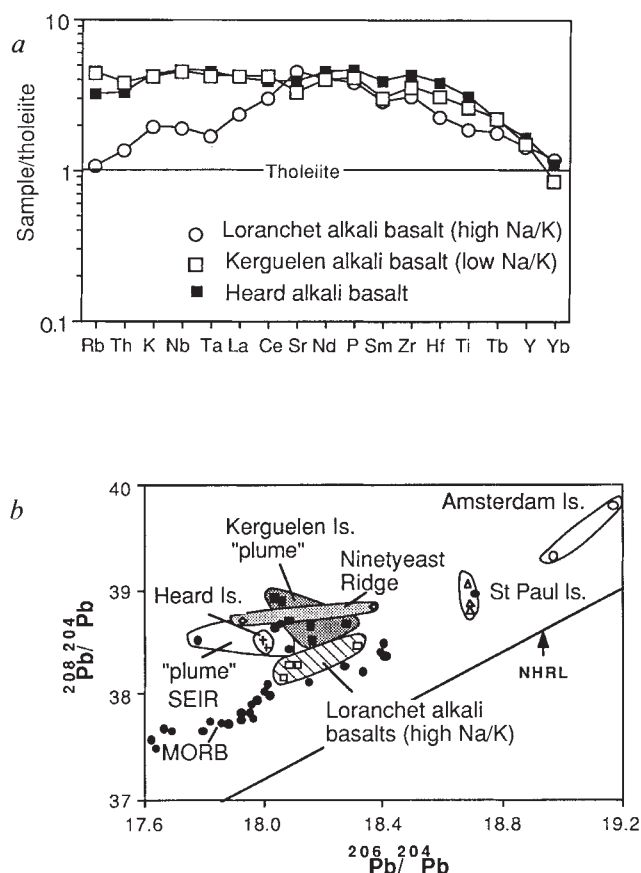


Fig. 2 *a*, Incompatible element abundances of high Na/K (BM64886) and low Na/K (1967, P8[5]) alkali basalts from Kerguelen Island, and Heard Island basalt (BM64992), all normalized to tholeiite (BM64947) from the Loranchet Peninsula. Element incompatibility increases from right to left. Note the strong similarity between patterns of the Heard Island basalt and the low Na/K basalt from Kerguelen Island, consistent with derivation from similar sources. The negative slope between Zr and Yb suggests the involvement of residual garnet in the formation of all these basalt types. *b*, Plots of $^{143}\text{Nd}/^{144}\text{Nd}$ against $^{87}\text{Sr}/^{86}\text{Sr}$ and (inset), $1/\text{Sr}$ against $^{87}\text{Sr}/^{86}\text{Sr}$. Data sources: Heard Island, this study; Kerguelen Island, refs 28, 29 and this study; Foch Island, ref. 18; Ninetyeast Ridge, ref. 7; St Paul Island, refs 19, 21; SEIR MORB, refs 19–21. *c*, Plot of $^{208}\text{Pb}/^{204}\text{Pb}$ against $^{206}\text{Pb}/^{204}\text{Pb}$. Data sources: Heard Island, this study; Kerguelen Island, ref. 28 and this study; Ninetyeast Ridge, ref. 30; St Paul Island, Refs 19, 21, 30; Amsterdam Island, ref. 30; SEIR MORB, refs 19–21. NHRL denotes the Northern Hemisphere reference line, defined by the data array for Northern Hemisphere MORBs and ocean island basalts²⁴.

and to other within-plate oceanic islands with highly enriched source characteristics^{22,23}.

The data show that most Kerguelen and Heard Island basalts have strong DUPAL characteristics²⁴, which is to say that they have high $^{87}\text{Sr}/^{86}\text{Sr}$, $^{207}\text{Pb}/^{206}\text{Pb}$ and $^{208}\text{Pb}/^{206}\text{Pb}$ ratios relative to most oceanic basalts. Basalts from the Ninetyeast Ridge have similar (DUPAL) isotopic compositions to those of most Kerguelen and Heard Island basalts (Fig. 2). As the Ninetyeast Ridge was emplaced on very young oceanic crust, this indicates that the DUPAL component in Kerguelen and Heard Island basalts is not due to contamination by the old, Cretaceous plateau lithosphere but resides within either the plume or the asthenosphere. Furthermore, the longevity of the DUPAL component within this single hotspot system strongly suggests that it is an intrinsic feature of the Kerguelen/Heard mantle plume.

A second aspect of Kerguelen Island basalts is the time evolution of their composition, manifest in terms of both isotopic and trace element ratios. On Kerguelen the transition from the older, predominantly tholeiitic volcanism to younger (type 3) alkali basalt volcanism is marked by an increase in the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio and a decrease in the $^{143}\text{Nd}/^{144}\text{Nd}$ ratio of erupted basalts. This is the opposite of what is observed in basalts from individual islands related to the Hawaiian plume². This observation, coupled with the slow motion of the Antarctic Plate relative to the Kerguelen hotspot (0.6 cm yr^{-1})²⁵, indicates that the evolution of the island cannot be interpreted simply in terms of the waning influence of a mantle plume (unlike the Hawaiian chain, where the Pacific Plate is moving an order of magnitude faster). Nor is there any reason to suppose that the physicochemical structure of this major plume has changed rapidly enough to account for the observed magmatic evolution. We suspect that the principal cause for the compositional change of Kerguelen Island basalts was the migration of the SEIR away from the vicinity of the plume. The lower $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of the early Kerguelen Island tholeiites could be interpreted in terms of plume/MORB asthenosphere interaction. An alternative, but

not mutually exclusive, explanation is that the plume-derived melts have suffered contamination by lithospheric magmas (see below). The voluminous nature of Kerguelen Island tholeiites and their lower abundances of the incompatible elements relative to younger low-Na/K alkali basalts, suggest that they originate from larger-degree melts of the mantle plume, which may again involve the influence of the SEIR. Extensional tectonics and the presence of thinner lithosphere would favour a large depth-melting interval for an upwelling adiabatic plume. This depth-melting interval would have decreased as the plume became increasingly separated from the SEIR, thus providing an explanation for the transition from tholeiitic to alkalic basalts (type 3) observed on Kerguelen Island.

The incompatible element abundances and isotopic ratios of the high-Na/K (type 2) alkali basalts suggest that they are the product of low degrees of melting of a source that is less enriched than that responsible for other basalts from Kerguelen. One explanation is that they derive from MORB asthenosphere which became entrained at the edge of the plume, thereby undergoing forced convection, melting and subsequent mixing with plume-derived melts. Alternatively, they may represent small-degree melts of the lithosphere. This is not inconsistent with our preliminary data, which indicate that tholeiitic basement samples dredged from the northern part of the Kerguelen Plateau (by the *Marion Dufresne* in 1986¹¹) have ratios of immobile incompatible elements (for example, $\text{Zr}/\text{Nb} = 14$) similar to those of the high-Na/K alkali basalts from Kerguelen Island. The early near-ridge setting of Kerguelen Island probably facilitated large-degree melting of the plume, as manifested by the shield tholeiites. Hence, pooling of high-temperature magmas at the base of the lithosphere may have been sufficient to cause small amounts of melting of this carapace, as has been suggested for Hawaii^{2–4}. The slow movement of the Antarctic Plate may make the Kerguelen Plateau lithosphere particularly vulnerable to melting by a plume. Contamination of plume-derived magmas by lithosphere-derived melts may partly account for the inverse

correlation between Sr content and $^{87}\text{Sr}/^{86}\text{Sr}$ ratio observed for the Loranchet alkali and tholeiitic basalts and would suggest an alternative explanation for the less enriched source characteristics of early Kerguelen tholeiites.

Seasat gravity maps show that part of the SEIR segment, lying between the Kerguelen Plateau and Broken Ridge and encompassing Amsterdam and St Paul Islands, is characterized by a strong geoid high⁸. Most basalt samples from this ridge segment (the 'SEIR plume' segment) have Sr, Nd and Pb isotopic compositions that overlap with those from Kerguelen and Heard Islands, although some show affinities with samples from the nearby Amsterdam and St Paul Islands (Figs 1 and 2)¹⁹⁻²¹. Basalts from the latter two islands do not have the distinctive Kerguelen geochemical signature, thus suggesting that they derive from a separate plume mantle source. Is the occurrence of basalts on the SEIR with Kerguelen/Heard DUPAL compositional characteristics evidence for the lateral flow of the plume from beneath the Kerguelen Plateau to the ridge axis? Perhaps the presence of progressively thinner lithosphere between the Kerguelen Plateau and the ridge axis channels plume flow upward towards the SEIR. Morgan²⁶ has postulated such channelled sub-lithospheric plume flow models. The large variability in isotopic ratios exhibited by basalts from this SEIR segment can be explained by mixing of the DUPAL plume with MORB asthenosphere. It is well established that Indian Ocean MORBs are compositionally distinct from Atlantic and Pacific MORBs, having higher $^{208}\text{Pb}/^{206}\text{Pb}$ and $^{207}\text{Pb}/^{206}\text{Pb}$ ratios¹⁹⁻²¹. We suggest that this distinct compositional 'mantle domain' could reflect wholesale contamination of the Indian Ocean asthenosphere by the Kerguelen/Heard DUPAL plume over the past 100 Myr or more, although a high $^{206}\text{Pb}/^{204}\text{Pb}$ component is also required to explain the mixing vectors in the Pb isotope plots (Fig. 2c). Such a component could be provided by hotspot material from Amsterdam and St Paul Islands.

We thank Alan Woolley and Valerie Jones of the British Museum (Natural History) for supplying samples. Nick Marsh carried out the XRF analyses. We appreciate discussion with Hubert Whitechurch, James Sevigny, Vincent Salters and Mike Norry. An earlier draft of this paper was substantially improved by the constructive criticisms of Hubert Whitechurch and W. M. White. This work was funded by the Natural Environment Research Council.

Received 12 July; accepted 6 October 1988.

1. Morgan, W. J. *Bull. Am. As. Petrol. Geol.* **56**, 203-213 (1972).
2. Chen, C.-Y. & Frey, F. A. *J. geophys. Res.* **90**, 8743-8768 (1985).
3. Gurriet, P. *Earth planet. Sci. Lett.* **82**, 153-158 (1987).
4. Ribe, N. M. *Earth planet. Sci. Lett.* **88**, 37-46 (1988).
5. Geist, D. J., White, W. M. & McBirney, A. R. *Nature* **333**, 657-660 (1988).
6. Morgan, W. J. in *The Sea* Vol. 7, (ed. Emiliani, E.) 443-487 (Wiley, New York, 1981).
7. Mahoney, J. J., Macdougall, J. D., Lugmair, G. W. & Gopalan, K. *Nature* **303**, 385-389 (1983).
8. Haxby, W. F. *Gravity Field of the World's Oceans* map (Natn. geophys. Data Centre, Boulder, Colorado, 1987).
9. Dietz, R. S. & Holden, J. C. *J. geophys. Res.* **75**, 4939-4966 (1970).
10. Leg 120 Shipboard scientific party *Nature* **334**, 385-386 (1988).
11. Leclaire, L. et al. *Geo-Marine Lett.* **7**, 169-176 (1987).
12. Munsch, M. & Schlich, R. *Mar. Geol.* **76**, 131-152 (1987).
13. Watkins, N. D., Gunn, B. M., Nougier, J. & Baksi, A. K. *Bull. Geol. Soc. Am.* **85**, 201-212 (1974).
14. Giret, A. & Lameyre, J. in *Antarctic Earth Science* (eds Oliver, R. L., James, P. R. & Jago, J. B.) 646-651 (Cambridge Univ. Press, 1983).
15. Quilty, P. G., Shafik, S., McMinn, A., Brady, H. & Clarke, I. in *Antarctic Earth Science* (eds Oliver, R. L., James, P. R. & Jago, J. B.) 636-639 (Cambridge Univ. Press, 1983).
16. Clarke, I., McDougall, I. & Whitford, D. J. in *Antarctic Earth Science* (eds Oliver, R. L., James, P. R. & Jago, J. B.) 631-635 (Cambridge Univ. Press, 1983).
17. Renard, A. *The Voyage of HMS Challenger* Vol. 2 (1889).
18. White, W. M. & Hofmann, A. W. *Nature* **296**, 821-825 (1982).
19. Dosso, L., Bougault, H., Beuzart, P., Calvez, J.-C. & Joron, J.-L. *Earth planet. Sci. Lett.* **88**, 47-59 (1988).
20. Hamelin, B., Dupre, B. & Allégre, C. J. *Earth planet. Sci. Lett.* **76**, 288-298 (1986).
21. Michard, A., Montigny, R. & Schlich, R. *Earth planet. Sci. Lett.* **78**, 104-114 (1986).
22. Palacz, Z. A. & Saunders, A. D. *Earth planet. Sci. Lett.* **79**, 270-280 (1986).
23. Weaver, B. L., Wood, D. A., Tarney, J. & Joron, J.-L. *Geology* **14**, 275-278 (1986).
24. Hart, S. R. *Nature* **309**, 753-757 (1984).
25. Minster, J. & Jordan T. H. *J. geophys. Res.* **83**, 5331-5354 (1978).
26. Morgan, W. J. *J. geophys. Res.* **83**, 5355-5360 (1978).
27. McKenzie, D. P. & Schlater, J. G. *Geophys. J. R. astr. Soc.* **25**, 437-528 (1971).
28. Dosso, L. et al. *Earth planet. Sci. Lett.* **43**, 46-60 (1979).
29. Dosso, L. & Murthy, V. R. *Earth planet. Sci. Lett.* **48**, 268-276 (1980).
30. Dupre, B. & Allégre, C. J. *Nature* **303**, 142-146 (1983).

Familial predisposition to Wilms' tumour does not map to the short arm of chromosome 11

Paul Grundy*, Alex Koufos*, Kenneth Morgan†, Frederick P. Li‡, Anna T. Meadows§ & Webster K. Cavenee*

*Ludwig Institute for Cancer Research, Royal Victoria Hospital, Montreal H3A 1A1, Canada

†Department of Epidemiology and Biostatistics, McGill University, Montreal H3A 1A2, Canada

‡Clinical Epidemiology Branch, National Cancer Institute and Division of Biostatistics and Epidemiology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA

§Division of Oncology, Children's Hospital of Philadelphia and University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA

Wilms' tumour of the kidney usually occurs sporadically, but can also segregate as an autosomal dominant trait with incomplete penetrance^{1,2}. Patients with the WAGR syndrome of aniridia, genitourinary anomalies, mental retardation and high risk of Wilms' tumour have overlapping deletions of chromosome 11p13 (ref. 3) which has suggested a possible location for a Wilms' tumour locus. Moreover, many sporadic tumours have lost a portion of chromosome 11p (refs 4-8). A second locus at 11p15 is implicated by association of the tumour with the Wiedemann-Beckwith syndrome (refs 9-11) and by tumour-specific losses of chromosome 11 confined to 11p15 (ref. 12). Here we report a multipoint linkage analysis of a family segregating for Wilms' tumour, using polymorphic DNA markers mapped to chromosome 11p. The results exclude the predisposing mutation from both locations. In a second family, the 11p15 alleles lost in the tumour were derived from the affected parent, thus precluding this region as the location of the inherited mutation. These findings imply an aetiological heterogeneity for Wilms' tumour and raise questions concerning the general applicability of the carcinogenesis model that has been useful in the understanding of retinoblastoma^{13,14}.

The pedigree of a family segregating for Wilms' tumour¹⁵ is shown in Fig. 1. There was no family history of consanguinity or of congenital anomalies such as aniridia, Wiedemann-Beckwith syndrome, hemihypertrophy or genitourinary malformations. Cytogenetic analyses of 17 family members as well as high-resolution chromosome banding (850-band level) in selected cases revealed no abnormality of constitutional chromosomes. As the incidence of Wilms' tumour is approximately 1 in 10,000 children¹⁶, this familial cluster is unlikely to be due to chance. The putative defect conferring a predisposition to Wilms' tumour in this family will hereafter be referred to as the WT locus, as distinct from the WAGR complex. The disease was considered to segregate as an autosomal dominant trait with reduced penetrance on the basis of two independent analyses of Wilms' tumour families^{1,2}, which included examples of skipped generations and affected cousins related by nonpenetrant presumed carriers, such as are apparent in this family.

Segregation analysis of seven DNA restriction fragment length polymorphisms (RFLP) at loci that span chromosome 11p was used to determine the position of the WT locus on chromosome 11p. The 7 marker loci in order were: centromere-CAT-{D11S16-FSHB}-CALCA-PTH-HBG2-Ha-ras1-11pter (refs 17 and 18; representing genes encoding the proteins catalase, follicle stimulating hormone, β -subunit, calcitonin, parathyroid hormone and haemoglobin- γ G). Importantly, analyses of the 11p deletion chromosomes of WAGR syndrome patients have shown that CAT and FSHB flank the WAGR complex in 11p13 (refs 19 and 20) and that D11S16 is tightly linked to CAT (ref. 21), although it appears closer to FSHB in physical mapping

Fig. 1 An abridged pedigree of a family with Wilms' tumour¹⁵. IV-9 and IV-14 were diagnosed as the family was under study. Individuals indicated were evaluated by physical examination, intravenous pyelogram and constitutional karyotyping as previously described¹⁵. DNA was isolated from peripheral blood of all individuals except I-2, III-2, III-4, III-16, IV-12 and IV-13. The genotype of I-2 was inferred at all loci on the basis of the segregation of the markers in his 12 offspring. Letters below each individual represent haplotypes of the chromosome 11p13 marker genotypes. *CAT*, *D11S16*, *FSHB* haplotypes are: A = 1,1,1; B = 1,3,1; C = 2,3,1; D = 1,3,2; E = 2,1,2; F = 1,2,1; G = 2,1,1; H = 2,2,1; I = 1,2,2; J = 2,3,2; K = 2,2,2; L = 1,1,2. Both II-1 and III-5 had a recombinant haplotype ($J^* = 2,3,2$ and $C^* = 2,3,1$ respectively). Methods as in the legend to Fig. 2.

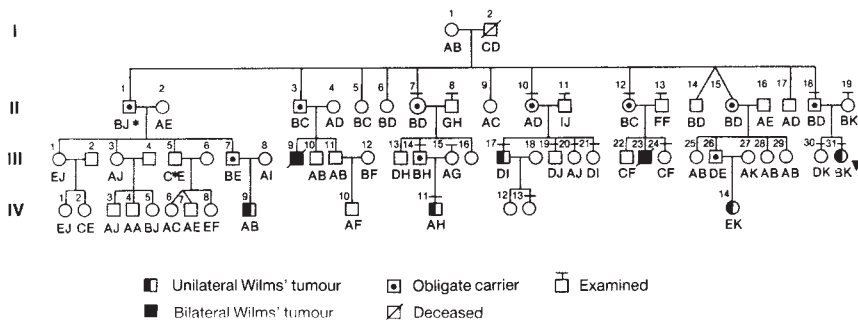
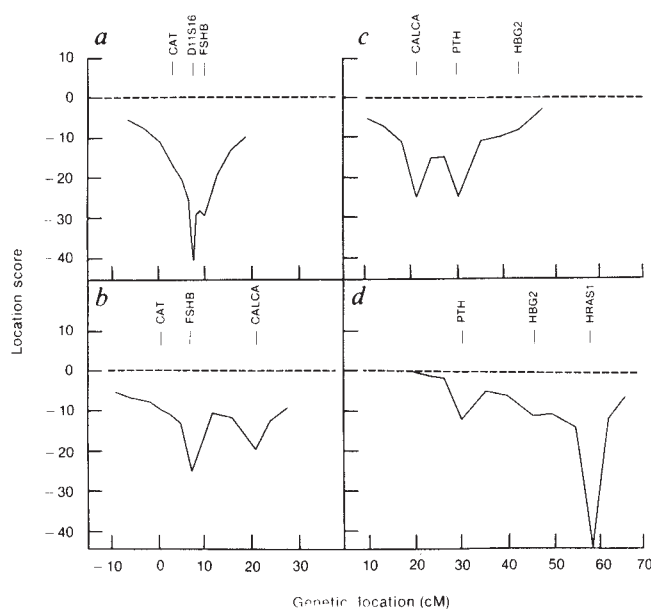


Fig. 2 Twice the natural logarithm of the likelihood ratio (location score) comparing the possible locations of the *WT* locus either on 11p or at a point infinitely removed from the group of fixed marker loci. *CAT* was designated map position 0.0; map distances for the other loci were obtained by converting the following recombination fractions using Haldane's mapping function²⁹: *CAT*(18%)-*CALCA*(8%)-*PTH*(12%)-*HBG2*(13%)-*Ha-ras1* (ref. 17). The 13% recombination rate between *HBG2* and *Ha-ras1* represents the sum of rates between *HBG2* and insulin (10%), and insulin and *Ha-ras1* (3%) (ref. 17). *FSHB* was assumed to be 6 cM distal to *CAT* (ref. 30). *D11S16* was placed between *CAT* and *FSHB*, 4 cM distal to *CAT* (refs 18 and 21). Location scores were computed using LINKMAP (ref. 23) with the following assumptions: autosomal dominant mode of inheritance with the disease penetrance in the heterozygote being 63% (ref. 1), and in the homozygote 100%; a population frequency of the *WT* gene of 0.0001, based on the incidence of Wilms' tumour; and a zero spontaneous mutation rate at the *WT* locus. Varying these parameters within reasonable limits had no significant effect on the results. Only one disease liability class was considered because 95% of Wilms' tumours occur by the age of 10 years³¹. The location of the *WT* locus was considered relative to overlapping subsets of the marker loci. *a*, *CAT*, *D11S16*, *FSHB*; *b*, *CAT*, *FSHB*, *CALCA*; *c*, *CALCA*, *PTH*, *HBG2*; *d*, *PTH*, *HBG2*, *Ha-ras1*.

Methods. DNA was digested with the restriction enzyme indicated below, electrophoresed through agarose and transferred to Nytran membranes (Schleicher and Schuell) as described³². Membranes were hybridized to probes labelled with [³²P]dCTP and washed under standard conditions. The probes (locus name) and restriction enzymes used were: pCATint800 (*CAT*), *TaqI* (ref. 33); p32-1(*D11S16*), *MspI* (ref. 34); pFSHB0.5 (*FSHB*), *HindIII* (ref. 35); pHC9 (*CALCA*), *TaqI* (ref. 36); p20.36(*PTH*), *TaqI* (ref. 37); JW151(*HBG2*), *HindIII* (ref. 38); pTBB-2 (*Ha-ras1*), *TaqI* (ref. 39). The six observed alleles of *Ha-ras1* were coded as a 4-allele locus⁴⁰.



studies¹⁸. The other four loci have been localized to the 11p15 region¹⁷.

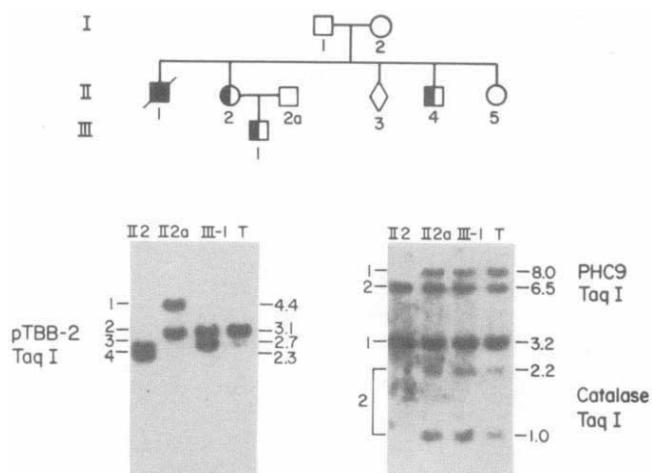
Initially each individual's haplotype for the 11p13 loci (*CAT*, *D11S16* and *FSHB*) was determined by inspection of segregation of the markers (Fig. 1). Surprisingly, four affected individuals (III-17, III-31, IV-11 and IV-14) do not share any 11p13 haplotype, suggesting that the *WT* locus cannot be located within the region defined by the flanking loci *CAT* and *FSHB*. Although II-1 had a recombinant haplotype, his affected descendant (IV-9) inherited the non-recombinant haplotype, thus precluding the possibility that this recombination event was related to the development of disease in this individual.

The likelihood of linkage of the *WT* locus to the 11p13 marker group as well as to the remainder of chromosome 11p was quantified by multipoint linkage analysis using the LINKMAP program^{22,23}. We compared the likelihood ratios for the possible location of the *WT* locus either on 11p or at a point infinitely removed from the group of linked markers, and twice the natural logarithm of the likelihood ratio (location score) is shown in Fig. 2a-d. The criterion used to exclude the *WT* locus from a given location was a location score of -9.2 or less, that is, odds

for a location of less than or equal to 1 in 100. Only three marker loci were treated at a time, but by considering the location of the *WT* locus relative to adjacent or overlapping sets of marker loci, we analysed the short arm of chromosome 11 from *CAT* to *Ha-ras1*. Location of the *WT* locus is excluded from 3 cM (centimorgan), centromeric to *CAT* to 5 cM telomeric to *Ha-ras1*, the most distally mapped locus on chromosome 11p. Similar results were obtained regardless of the location assumed for *D11S16* relative to *CAT* and *FSHB*, or when the distance between *CAT* and *FSHB* was varied between 1 and 10 cM (data not shown).

As a possible cytogenetically undetected interstitial translocation or rearrangement of the short arm of chromosome 11 could have occurred in this family, recombination frequencies between the marker loci were determined and were consistent with those obtained in families without disease¹⁷. In particular, *CAT* and *FSHB* showed significant evidence of linkage ($\theta = 0.075$, $Z_{\max} = 3.45$). Thus there was no evidence of a rearrangement of chromosome 11p. An unbalanced interstitial translocation of 11p13 including *CAT* and *FSHB*, which has been described in one Wilms' tumour family²⁴, is also unlikely, because each of the

Fig. 3 Pedigree of a second family segregating for Wilms' tumour (top). II-2 was the proband²⁵, III-1 subsequently developed a Wilms' tumour. Autoradiographs of Southern blots (bottom) of DNA from indicated family members, hybridized to pTBB-2 (Ha-ras1), pHC9 (CALCA) and pCAT int800 (CAT). *T* represents DNA from primary tumour tissue. Numbers on the left indicate the alleles. Numbers on the right are size markers in kilobases. Methods as in legend to Fig. 2.



five affected individuals with known genotypes were constitutionally heterozygous for at least one of these markers (Fig. 1). A small interstitial translocation or deletion involving only the region between the aniridia locus and *FSHB* (as no family member had aniridia) cannot be excluded, but this possibility is less likely, given the similar evidence against linkage in the accompanying paper.

The second family²⁵ that we studied is shown in Fig. 3. Notably, the grandmother (I-2) had hemihypertrophy, a condition often seen as part of Wiedemann-Beckwith syndrome. Since the initial report, the sole child of individual II-2 has developed a unilateral Wilms' tumour. Comparison of this child's (III-1) constitutional and tumour genotypes for RFLP loci mapped to chromosome 11p indicated a loss of constitutional heterozygosity at the most distal locus, Ha-ras1 (11p15.5), but maintenance of the constitutional genotype at the more proximal loci *CALCA* (11p15.4) and *CAT* (11p13) (Fig. 3). This is consistent with data from other Wilms' tumours (our unpublished data and ref. 12) that have identified losses of constitutional heterozygosity for the portion of 11p distal to the *WAGR* complex. As the parents of III-1 had four different Ha-ras1 alleles, the parental origin of the Ha-ras1 allele lost from the tumour could be unambiguously identified. As shown in Fig. 3, the 2.7-kilobase(kb) allele lost from the tumour (T) was inherited from the affected mother (II-2); these data are inconsistent with the unmasking of an inherited recessive defect at 11p15 in the genesis of the tumour. A similar situation has been described for multiple endocrine neoplasia type 2 (ref. 26).

Genetic linkage analysis should be able to detect the locus of the inherited germinal mutation that confers a predisposition to Wilms' tumour by any model of inheritance, including the two-hit model¹, the host-resistance model², or a model incorporating two-hits and sex-specific genomic imprinting²⁷.

That a gene within the *WAGR* complex at 11p13 is involved in the genesis of Wilms' tumour, at least in those children with the *WAGR* syndrome, seems indisputable, given the number of cases described, and that the introduction of a normal human chromosome 11 into a Wilms' tumour cell line controlled its tumorigenic expression²⁸. Similar arguments support the existence of a locus at 11p15 that is associated with the Wiedemann-Beckwith syndrome. But our findings strongly imply the existence of another locus that, when mutated, is capable of conferring an inherited predisposition to Wilms' tumour without other anomalies.

Several possible models of carcinogenesis could follow from our results. Wilms' tumours may arise as a result of inactivation of both alleles at a single locus, as has been suggested¹³, but different loci may be implicated in different tumours, even though the tumours are histologically and clinically indistin-

guishable. These loci might be involved in different steps of the same regulatory pathway, the abrogation of which could lead to the malignant state. Alternatively, the two hits could be at different loci, if one allele at a locus was already inactivated through a mechanism such as genomic imprinting. Finally, the development of Wilms' tumour may require more than two events, one each at the *WAGR* complex at the 11p15 locus and at what is so far an unidentified location. This seems least likely given the improbability of three events occurring in a single cell and the young age of presentation of Wilms' tumours.

We thank Dr Gail Bruns for collaboration in providing DNA from the family members and Drs R. A. Gravel, P. C. Watkins, S. Hoppener, H. Mayer, H. Kazazian and M. Wigler for DNA probes. We also thank A. Kalbakji for technical assistance and T. Genio for preparation of the manuscript.

Received 12 August; accepted 19 October 1988.

- Knudson, A. G. Jr & Strong, L. C. *J. natn. Cancer Inst.* **48**, 313-324 (1972).
- Matsunaga, E. *Hum. Genet.* **57**, 231-246 (1981).
- Riccardi, V. M., Sujansky, E., Smith, A. C. & Francke, U. *Pediatrics* **61**, 604-610 (1978).
- Koufos, A. *et al. Nature* **309**, 170-172 (1984).
- Orkin, S. H., Goldman, D. S. & Sallan, S. E. *Nature* **309**, 172-174 (1984).
- Reeve, A. E. *et al. Nature* **309**, 174-176 (1984).
- Feareon, E. R., Vogelstein, B. & Feinberg, A. P. *Nature* **309**, 176-178 (1984).
- Dao, D. D. *et al. Am. J. hum. Genet.* **41**, 202-217 (1987).
- Sotelo-Avila, C. & Gooch, W. M. *Perspect. pediat. Path.* **3**, 255-272 (1976).
- Waziri, M., Patil, S. R., Hanson, J. W. & Bartley, J. A. *J. Pediatr.* **102**, 873-876 (1983).
- Turleau, C. J. *et al. Hum. Genet.* **67**, 219-221 (1984).
- Raizis, A. M., Becroft, D. M., Shaw, R. L. & Reeve, A. E. *Hum. Genet.* **70**, 344-346 (1985).
- Knudson, A. G. Jr *Cancer Res.* **45**, 1437-1443 (1985).
- Cavenee, W. K. *et al. Nature* **305**, 779-784 (1983).
- Cordero, J. F., Li, F. P., Holmes, L. B. & Gerald, P. S. *Pediatrics* **66**, 716-719 (1980).
- Wilimas, J. in *Pediatrics* 1110-1113 (Appleton & Lange, Norwalk, 1987).
- Grzeschick, K. H. & Kazazian, H. H. *Cytogenet. Cell Genet.* **40**, 179-205 (1985).
- Glaser, T. M. *et al. Cytogenet. Cell Genet.* **46**, 620 (1987).
- Van Heyningen, V. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 8592-8596 (1985).
- Kazazian, H. H. & Junien, C. *Cytogenet. Cell Genet.* **46**, 188-212 (1987).
- Kidd, J. R., Castiglione, A. J., Pakstis, A. J. & Kidd, K. K. *Cytogenet. Cell Genet.* **45**, 63-64 (1987).
- Lathrop, G. M. & Lalouel, J. M. *Am. J. hum. Genet.* **36**, 460-465 (1984).
- Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3443-3446 (1984).
- Yunis, J. J. & Ramsay, N. K. C. *J. pediat.* **96**, 1027-1030 (1980).
- Meadows, A. T., Lichtenfeld, J. L. & Koop, C. E. *New Engl. J. Med.* **291**, 23-24 (1974).
- Mathew, C. G. P. *et al. Nature* **328**, 524-526 (1987).
- Wilkins, R. J. *Lancet* **i**, 329-331 (1988).
- Weissman, B. E. *et al. Science* **236**, 175-180 (1987).
- Haldane, J. B. S. *J. Genet.* **8**, 299-309 (1919).
- Glaser, T. *et al. Nature* **321**, 882-887 (1986).
- Breslow, N., Beckwith, J. B., Ciol, M. & Sharples, K. *Cancer Res.* **48**, 1653-1657 (1988).
- Koufos, A. *et al. Nature* **316**, 330-334 (1985).
- Quan, F., Korneluk, R. G., MacLeod, H. L., Tsui, L. C. & Gravel, R. A. *Nucleic Acids Res.* **13**, 8288 (1985).
- Feder, J. *et al. Am. J. hum. Genet.* **37**, 635-649 (1985).
- Watkins, P. C. *et al. DNA* **6**, 205-212 (1987).
- Hoppener, J. W. M. *et al. Hum. Genet.* **66**, 309-312 (1984).
- Schmidtke, J. *et al. Hum. Genet.* **67**, 428-431 (1984).
- Jeffreys, A. J. *Cell* **18**, 1-10 (1979).
- Barker, D. *et al. Molec. biol. Med.* **1**, 199-206 (1983).
- Ott, J. *Ann. hum. Genet.* **42**, 255-257 (1978).

Lack of linkage of familial Wilms' tumour to chromosomal band 11p13

Vicki Huff*, Duane A. Compton*, Lian-Yu Chao*,
Louise C. Strong†, Clementina F. Geiser‡
& Grady F. Saunders*§

* Department of Biochemistry and Molecular Biology, University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd, Houston, Texas 77030, USA

† Department of Experimental Pediatrics/Genetics, University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA

‡ Division of Hematology/Oncology, University of Texas Health Science Center, San Antonio, Texas 78284, USA

Wilms' tumour (WT), a paediatric renal neoplasm, affects approximately 1 in 10,000 children. One or both kidneys can be affected and 5–10% of tumours are bilateral¹. Most tumours occur sporadically; however, around 1% of the cases are familial, with siblings or cousins most often being affected². Familial cases are more frequently bilateral³, and familial and bilateral tumours are diagnosed at an earlier age¹. On the basis of these observations, it was proposed that the development of WT requires two mutations⁴. In most sporadic unilateral WT, both are somatic; in familial and bilateral tumours the first is thought to be germinal. Cytogenetic and molecular studies have demonstrated germinal mutations in WT/aniridia patients and somatic mutations in sporadic WT at chromosomal band 11p13. To investigate whether familial predisposition to WT is due to a germinal 11p13 mutation, we studied a WT family with seven DNA markers that span the 11p13 region. We found that familial WT predisposition was not genetically linked to any of the 11p13 markers. This suggests that the gene involved in familial WT predisposition is outside 11p13 and is distinct from the gene involved in tumorigenesis and in WT predisposition in WT/aniridia 11p13-deletion patients.

Cytogenetic observations of constitutional chromosome 11p13 deletions in patients with WT/aniridia and of tumour-specific 11p13 deletions in other Wilms' tumours that suggested that a locus at chromosomal band 11p13 was involved in tumorigenesis^{5–7}. Subsequent molecular analyses of tumours indicated that ~50% show loss of heterozygosity at 11p loci, suggesting that a mutation of one 11p13 allele and subsequent mutation or loss of the normal homologous allele are necessary for tumorigenesis^{8–13}.

Based on the above data, we expected that the familial predisposition to WT would be the result of a germinal mutation at 11p13. To investigate this, we studied a large WT family whose pedigree is shown in Fig. 1. Six children from three collateral sibships were diagnosed as having WT. The five cases for which histological data were available had 'typical' triphasic WT with favourable histology and subcapsular nephroblastomatosis. The occurrence of bilateral tumours, the early diagnosis, and the variable penetrance are consistent with observations of other familial WT cases². No congenital anomalies, hemihypertrophy, or visceromegaly were reported in the family, and cytogenetic evaluation did not reveal any detectable karyotypic aberration. DNA from the individuals whose identification number is underlined in the pedigree was typed for restriction fragment length polymorphisms (RFLPs) at loci in both 11p13 and in 11p15. Polymorphic loci at 11p15 were analysed with the DNA probes for H-ras¹⁴, insulin¹⁵ and parathyroid hormone¹⁶. Twelve 11p13 RFLPs were detected by seven chromosome band 11p13-specific DNA probes (Table 1). The location of these probes within band 11p13 is shown in Fig. 2. As indicated, the seven probes detect five different *NotI* restriction fragments that span the 11p13 region. Analysis of DNA from WT/aniridia patients with constitutional 11p13 deletions has defined an area between p8

Table 1 Chromosome 11p13 DNA probes and restriction endonucleases used to detect independent RFLPs for linkage analyses

Locus or haplotype	DNA probe	Restriction endonucleases	Number of alleles
FSHB/D11S151	FSHB-1.1	<i>HindIII</i>	2
	p56H2.4	<i>EcoRI</i>	2
	p56H2.4	<i>PstI</i>	2
D11S16 p5/p60	p32-1	<i>MspI</i>	3
	p5BE1.2	<i>HaeIII</i>	2
	p5S1.6	<i>PvuII</i>	2
	p60H1.4	<i>TaqI</i>	2
p8	p8B1.25	<i>HindIII</i>	2
	p8B1.25	<i>BglII</i>	2
	pCAT41	<i>KpnI</i>	2
CAT	pCAT41	<i>AvaII</i>	2
	pCAT41	<i>HaeIII</i>	2

p56, p5, p60 and p8 are unique-sequence DNA plasmid subclones from lambda phage genomic clones of the same number^{17,29}. FSHB-1.1 (ref. 30) and pCAT41 (ref. 31) are DNA clones for the genes for the β subunit of follicle-stimulating hormone and catalase, respectively. p32-1 is an anonymous DNA clone that maps to 11p13 (refs 18, 32).

and p5/p60 which is involved in the development of WT¹⁷. The probes used in this study flank this region.

Overall, there was no evidence for recombination between 11p13 markers in this pedigree; 11p13 loci segregated as a unit (data not shown), which is consistent with estimates of genetic distances in this region (Fig. 2; ref. 18). RFLP data obtained by probes for the same *NotI* restriction fragment were grouped together into haplotypes to make the genotype data more informative. This resulted in four haplotypes (FSHB/D11S151, p5/p60, p8 and CAT) and a single RFLP (D11S16). Linkage analysis using the 11p13 markers separately and in haplotype combinations ruled out linkage of any 11p13 marker to familial WT; lod scores of -5.17 to -2.84 were obtained at recombination fraction $\theta = 0.00$ (Table 2). As a lod score of -2 is considered to be significant evidence against linkage¹⁹, these data indicate that the familial predisposition to WT in this pedigree does not map to chromosomal band 11p13.

Wilms' tumour is occasionally observed in individuals with Beckwith-Wiedemann syndrome, which has been associated with genetic alterations at 11p15²⁰. Because of this, linkage of the WT predisposition in this family to 11p15 markers was also investigated. A lod score of -2 at $\theta = 0.00$ was obtained for the HRAS1 locus (data not shown), indicating that the gene for WT predisposition does not map to chromosomal band 11p15.

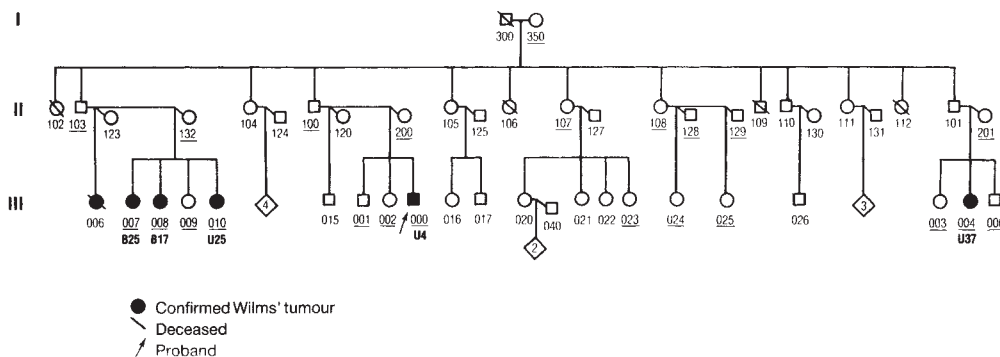
Lack of linkage of 11p13 markers to familial WT is unexpected in the face of considerable evidence that genetic alterations at 11p13 are important in the development of Wilms' tumour. Individuals with aniridia and a constitutional 11p13 deletion are at a greatly increased risk of WT (ref. 21). In addition, interstitial translocations involving 11p13 have been reported in families in which individuals with WT/aniridia inherited an 11p13 deletion chromosome through a balanced, unaffected

Table 2 Lod scores at various recombination fractions (θ) for linkage between WT and chromosome 11p13 markers

	0.00	0.05	0.10	0.15	0.20	0.25
FSHB/D11S151	-5.17	-0.72	-0.25	-0.03	+0.08	+0.13
D11S16	-2.89	-0.54	-0.31	-0.20	-0.14	-0.10
p5/p60	-5.05	-1.44	-0.93	-0.66	-0.48	-0.34
p8	-2.84	-0.48	-0.26	-0.17	-0.12	-0.09
CAT	-3.75	-1.03	-0.62	-0.39	-0.23	-0.13

Likelihood analyses were performed using the LIPED program³³ assuming a penetrance of 0.63 for the WT gene, a gene frequency of 0.0001 and equal male and female recombination fractions. Varying the penetrance function did not appreciably alter the resultant lod scores.

Fig. 1 A pedigree in which Wilms' tumour (solid symbol) is observed in six children in the third generation. DNA from lymphoblastoid cell lines or peripheral lymphocytes was isolated from individuals whose identification number is underlined and RFLP data were obtained. B and U designations below the affected individuals' identification numbers indicate bilateral or unilateral tumours, respectively, and the age at diagnosis in months. Squares, males; circles, females; diamonds, child(ren) of unspecified sex.



carrier²²⁻²⁴. Thus, germinal deletions at 11p13 can predispose individuals to WT. It is interesting that the clinical characteristics of Wilms' tumours, including age at diagnosis, frequency of bilaterality and associated nephroblastomatosis, are similar in this family and in WT/aniridia patients²⁵. Tumour-specific loss of heterozygosity at 11p has been previously documented in ~50% of sporadic tumours⁸⁻¹³, further implicating chromosome 11 in the development of WT.

Our data indicate that the predisposing gene in this family is different from the gene important in WT/aniridia, in Beckwith-Wiedemann syndrome and in WT tumorigenesis. The nature of the interaction, if any, between these two types of genes *vis-à-vis* the development of tumours in predisposed individuals is unknown. The predisposing gene may lead to tumour formation independently of a 11p13 locus. Alternatively, tumour-specific genetic alterations at 11p13 may be required in addition to the predisposing mutation. The study of tumour-specific alterations at 11p13 in familial WT patients might be informative in this respect. Unfortunately, we have not had access to any tumour tissue.

In an alternative to the two-locus model, there would be a lack of linkage of the predisposing WT gene to 11p13 if the predisposition was a premutation that was itself unstable and could cause instability in its normal homologue. If such a premutation at one 11p13 allele resulted in an overt mutation at that allele in one family member, but at the homologous allele in a second family member, we would not have observed linkage between 11p13 markers and WT in our multiple sibship pedigree. Similarly, linkage would not be observed if a highly localized double recombination or gene conversion event occurred. However, there is no evidence of instability at 11p13 in this family; no recombination between multiple 11p13 markers was observed.

What is evident from these studies is that familial predisposition to WT in this family does not co-segregate with 11p13 markers. These data indicate that, barring the kind of unusual

mechanisms discussed above, the gene for familial predisposition to WT is distinct from the 11p13 gene that is implicated in WT/aniridia and tumorigenesis. A distinction between a predisposing gene and a tumorigenesis gene has been made for multiple endocrine neoplasia type 2 (refs 26-28). For Wilms' tumour the relationship between the two genes and their role, separately or jointly, in the development of sporadic and familial tumours is unknown. Nor has the location and nature of the predisposing gene been determined. Such information, together with a knowledge of the nature and function of the 11p13 gene, will be instrumental in understanding the biology of predisposition to Wilms' tumour.

We thank the family members and the following physicians who provided family samples for study: Dr Paul Zeltzer, Children's Hospital of Los Angeles, California and Dr C. E. Krill Jr, Children's Hospital, Akron, Ohio. We also thank Drs P. C. Watkins (Integrated Genetics), R. A. Gravel, R. A. Weinberg, G. I. Bell and H. Kronenberg for providing us with plasmids pFSHB-1.1, pCAT41, pEJ, pHINS300 and pPTHm122, respectively. We thank Dr W. W. Nichols for establishing lymphoblastoid cell lines and Dr W. T. Schroeder for her interest in this study. This work was supported by grants from the NIH and the Robert A. Welch Foundation.

Received 12 August; accepted 30 September 1988.

- Breslow, N. E. & Beckwith, J. B. *J. natn. Cancer Inst.* **68**, 429-436 (1982).
- Matsunaga, E. *Hum. Genet.* **57**, 231-246 (1981).
- Strong, L. C. in *Clinical Pediatric Oncology* 3rd edn (eds Sutow, W. W., Vietti, T. J. & Fernbach, D. J.) 14-41 (Mosby, St Louis, 1984).
- Knudson, A. G. & Strong, L. C. *J. natn. Cancer Inst.* **48**, 313-324 (1972).
- Riccardi, V. M. *et al. Pediatrics* **61**, 604-610 (1978).
- Kaneko, Y., Egues, M. C. & Rowley, J. D. *Cancer Res.* **41**, 4577-4578 (1981).
- Slater, R. M. & de Kraker, J. *Cancer Genet. Cytogenet.* **5**, 237-245 (1982).
- Fearon, E. R., Vogelstein, B. & Feinberg, A. P. *Nature* **309**, 176-178 (1984).
- Koufos, A. *et al. Nature* **309**, 170-172 (1984).
- Orkin, S. H., Goldman, D. S. & Sallan, S. E. *Nature* **309**, 172-174 (1984).
- Reeve, A. F. *et al. Nature* **309**, 174-176 (1984).
- Raizis, A. M., Becroft, D. M., Shaw, R. L. & Reeve, A. E. *Hum. Genet.* **70**, 344-346 (1985).
- Dao, D. D. *et al. Am. J. hum. Genet.* **41**, 202-217 (1987).
- Shih, C. & Weinberg, R. A. *Cell* **29**, 161-169.
- Bell, G. I., Karam, J. H. & Rutter, W. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5759-5763 (1981).
- Hendy, G. N., Kronenberg, H. M., Potts, J. T. Jr & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7365-7369 (1981).
- Compton, D. A. *et al. Cell* (in the press).
- Kidd, J. R., Castiglione, C. M., Pakstis, A. J. & Kidd, K. K. *Cytogenet. Cell Genet.* **45**, 63-64 (1987).
- Ott, J. *Analysis of Human Genetic Linkage* (Johns Hopkins University Press, Baltimore, 1985).
- Koufos, A. *et al. Nature* **316**, 330-334 (1985).
- Narahara, K. *et al. Hum. Genet.* **66**, 181-185 (1984).
- Yunis, J. J. & Ramsay, N. K. C. *Pediatrics* **96**, 1027-1030 (1980).
- Koussief, B. G. *J. Pediatr.* **98**, 676-677 (1981).
- Nakagome, Y. & Nagahara, N. *Hum. Genet.* **70**, 289 (1985).
- Breslow, N., Beckwith, J. B., Ciol, M. & Sharples, K. *Cancer Res.* **48**, 1653-1657 (1988).
- Mathew, C. G. P. *et al. Nature* **328**, 524-526 (1987).
- Mathew, C. G. P. *et al. Nature* **328**, 527-528 (1987).
- Simpson, N. E. *et al. Nature* **328**, 528-530 (1987).
- Huif, V. *et al. Nucleic Acids Res.* **15**, 7651 (1987).
- Watkins, P. C. *et al. DNA* **6**, 205-212 (1987).
- Korneluk, R. G. *et al. J. biol. Chem.* **22**, 13819-13823 (1984).
- Feder, J. *et al. Am. J. hum. Genet.* **37**, 635-649 (1985).
- Ott, J. *Am. J. hum. Genet.* **26**, 588-597 (1974).

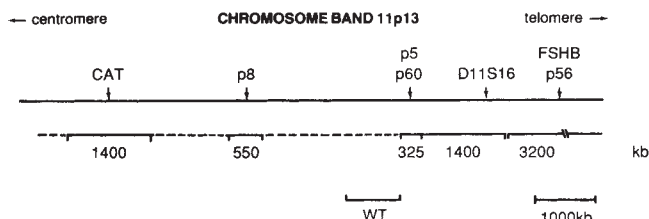


Fig. 2 Schematic diagram of chromosomal band 11p13 showing the location of the DNA probes used for linkage analysis relative to the inferred location of the locus for Wilms' tumour (WT). *NotI* restriction fragments are indicated by solid lines with size in kilobases noted underneath. Physical mapping of 11p13 indicates that the distance between D11S16 and CAT is ~5,000 kb¹⁷. The point estimate of θ between D11S16 and CAT is 0.00 (ref. 18).

Temporal integration by a slowly inactivating K⁺ current in hippocampal neurons

Johan F. Storm

Department of Neurobiology and Behavior and the Howard Hughes Medical Institute, State University of New York at Stony Brook, New York 11794, USA and Institute of Neurophysiology, Karl Johans gt. 47, 0162 Oslo 1, Norway

A central aspect of neuronal function is how each nerve cell translates synaptic input into a sequence of action potentials that carry information along the axon, coded as spike frequency. When transduction from a graded depolarizing input to spikes is studied by injecting a depolarizing current, there is often a remarkably long delay to the first action potential, both in mammalian and molluscan neurons¹⁻⁶. Here, I report that the delayed excitation in rat hippocampal neurons is due to a slowly inactivating potassium current, I_D . I_D co-exists with other voltage-gated K⁺ currents, including a fast A current and a slow delayed rectifier current. As I_D activates in the subthreshold range, and takes tens of seconds to recover from inactivation, it enables the cell to

integrate separate depolarizing inputs over long times. I_D also makes the encoding properties of the cell exceedingly sensitive to the prevailing membrane potential.

Figure 1a shows the response of a hippocampal CA1 pyramidal neuron to a current injection, with a typical long delay to the first action potential, during which the cell depolarized slowly, in a ramp-like way. The delay was much longer than the passive membrane time constant (≤ 20 ms). Furthermore, the delay increased substantially when the cell was hyperpolarized. Conversely, when the cell was depolarized to -64 mV, the delay decreased to less than 50 ms. It could also be reduced by increasing the pulse strength¹. In hyperpolarized cells, the delay could last for 10 s and the firing often accelerated before reaching a steady rate (Fig. 1b), that is, opposite behaviour to the adaptation seen at less polarized membrane potentials¹.

As shown by the voltage-clamp experiments in Fig. 1c, the delayed depolarization was associated with the induction of a rapidly activating, slowly inactivating outward current, called I_D (delay current). It is probably a voltage-gated K⁺ current as it was sensitive to the K⁺ gradient across the membrane but not to the Cl⁻ gradient. It also persisted when Ca²⁺ channels were blocked by cadmium, so it is not a Ca²⁺-activated current. Both the outward current and the underlying conductance were blocked by very low concentrations (30 μ M or so) of 4-aminopyridine (4-AP), but not by 25 mM tetraethyl ammonium (TEA). In these properties I_D resembles some other highly 4-AP-sensitive K⁺ currents⁷, as in axons⁸ and sensory ganglion cells⁹.

Present address: Institute of Neurophysiology, University of Oslo, Karl Johans gt. 47, 0162 Oslo 1, Norway.

Fig. 1 Delayed excitation in hippocampal neurons is due to a slowly inactivating K⁺ current, I_D . *a*, Response of a cell to depolarizing current. There was a long delay to the first action potential, during which the cell depolarized slowly, forming a ramp (dotted line). Resting potential -72 mV. *b*, The maximal delay exceeded 10 s in this hyperpolarized cell. *c*, Comparison of current-clamp and single-electrode voltage-clamp records. A fast activating, slowly inactivating outward current, I_D , underlay the slow ramp. *d*, I_D was blocked by 4-AP (30 μ M). Brief (40 ms) negative voltage steps were used to show the increase in membrane conductance when I_D was activated, and the decrease in conductance when I_D was blocked.

Methods. Transverse hippocampal slices from adult rats were prepared as described^{1,14}. The slices were kept between artificial cerebrospinal fluid (a.c.s.f.) and humid gas (95% O₂, 5% CO₂) at 34–36 °C (Figs 1a and 3a,d), or submerged in the a.c.s.f. at 26–32 °C (Figs 1b–d, 2a–c and 3b,c). The a.c.s.f. contained (in mM): NaCl, 122; NaHCO₃, 22–26; KCl, 3.0; CaCl₂, 2.0–2.5; MgCl₂, 1.0 and glucose, 10. CA1 pyramidal cells ($n=55$) were impaled with microelectrodes filled with 3 M KCl or 4 M potassium acetate, connected to a d.c. amplifier with a bridge circuit (Figs 1a and 2e); or a single-electrode voltage-clamp amplifier was used, for discontinuous current clamp (Fig. 1c, left) or voltage clamp (Figs 1c,d, 2a–c and 3b). The electrodes (14–40 M Ω) were coated to reduce capacitive coupling to the bath. The head-stage voltage was monitored to ensure complete settling during each switching cycle (frequency 4–10 kHz). Tetrodotoxin (0.5–1 μ M) was used to block spikes under voltage clamp. Cd²⁺ (200 μ M) was sometimes used to block synaptic activity and Ca²⁺-dependent currents (Figs 1b,d, 2a–c and 3b). All drugs were added to the perfusing a.c.s.f.

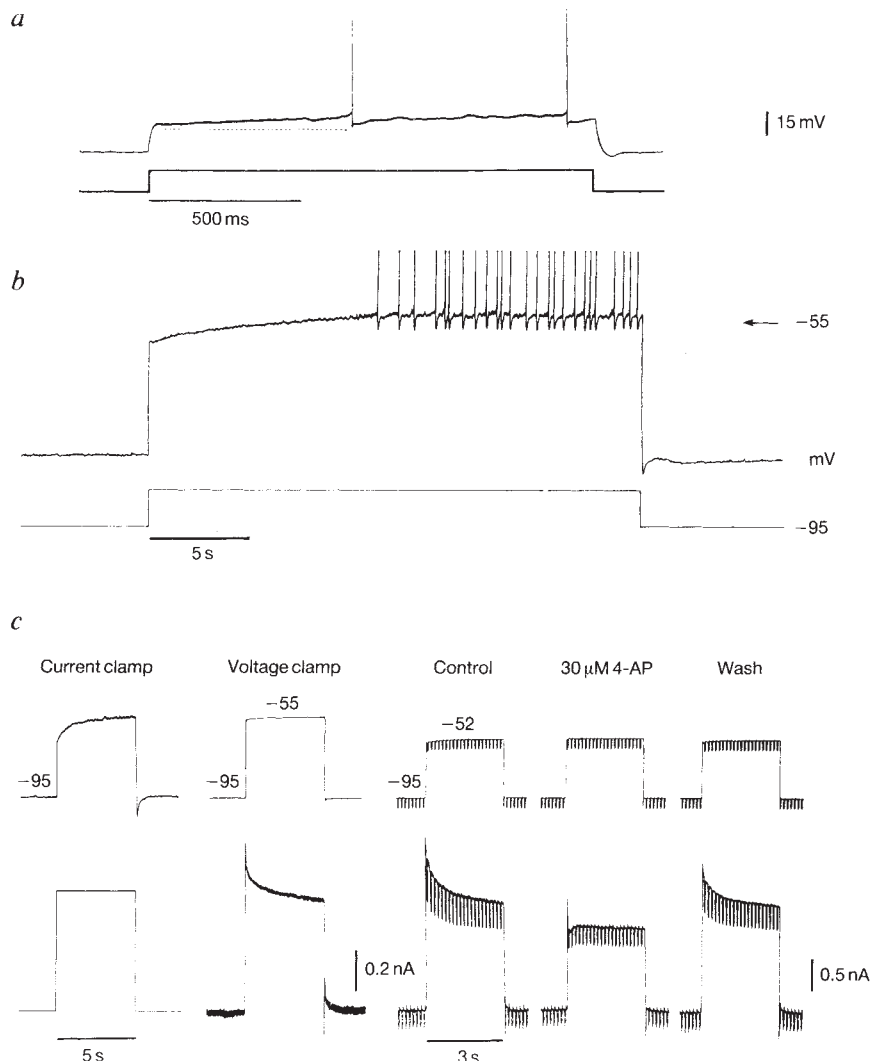


Fig. 2 The D-current differs from the A current. *a*, The transient K^+ current elicited by a step from -80 to -26 mV consisted of two components: a fast one, I_A , and a slow one, I_D . In contrast, a step from -60 to -26 mV elicited only I_A ; and $40 \mu\text{M}$ 4-AP blocked I_D but not I_A . *b*, Records of I_A and I_D showing voltage dependence of activation (left) and inactivation (right). The voltage steps are indicated below (mV). *c*, Plots of data from *b*: activation (filled circles) and inactivation (open circles). I_A was half-inactivated at -60 mV, I_D at -88 mV (dotted lines). *d*, Comparison between voltage-dependent K^+ currents in CA1 cells, indicating the kinetics (drawings of idealized currents), the approximate voltage ranges of inactivation and activation (shaded areas), typical voltage steps used to elicit the currents (solid lines), and effective blocking agents. Data for I_M are taken from ref. 14. *e*, Current-clamp records, showing that $40 \mu\text{M}$ 4-AP, which blocks I_D , also eliminated the long delay to the start of firing. The remaining short delay may be due to I_A . Response to three different current intensities (indicated below) are shown. Resting potential -74 mV. The spikes are truncated.

Methods. The currents of interest were isolated by adding Cd^{2+} ($200 \mu\text{M}$, *a-c*), to block Ca^{2+} -dependent currents. This may have caused a few millivolts positive shift in the activation and inactivation curves²⁰. In addition, $50 \mu\text{M}$ carbachol was used to block I_M and, sometimes, $1-3 \text{ mM}$ Cs^+ , to block I_Q , and $0.5-1 \text{ mM}$ 9-anthracene carboxylate to block the Cl^- -current $I_{\text{Cl(V)}}$ (ref. 16). The effectiveness 9-anthracene carboxylate was confirmed in voltage-clamp experiments like those in ref. 16, with Cs^+ -filled electrodes. In *a*, 7 mM TEA was used to reduce I_K . Dotted lines indicate leak current, as measured with hyperpolarizing steps. In *b*, I_A was isolated from I_D by adding $40 \mu\text{M}$ 4-AP. Although $40 \mu\text{M}$ 4-AP reduced the amplitude of I_A slightly (*a*), and speeded up its inactivation by about 20%, the voltage-dependence was not noticeably affected. *c*, I_A was measured 3 ms after the start of the test pulse, corresponding to the peak amplitude, and the leak current was subtracted. I_D was measured 50 ms after the pulse onset, when I_A has inactivated completely. The I_D amplitude was obtained by subtracting the current recorded in $40 \mu\text{M}$ 4-AP from control records (see *a*). The conductances were calculated on the basis of $E_K = -90$ mV. The abscissa in *c* indicates test-pulse voltages for the activation plot and pre-pulse voltages for the inactivation plot. As it was not always practical to use the test- and holding potentials which gave maximal conductance, different ordinate scales are used for inactivation (right) and activation (left) plots. In *e*, 4 mM manganese was added to both the control and the test medium, to block spontaneous synaptic potentials. Otherwise, the enhanced synaptic activity in 4-AP^{18,19} would contaminate the intrinsic firing pattern.

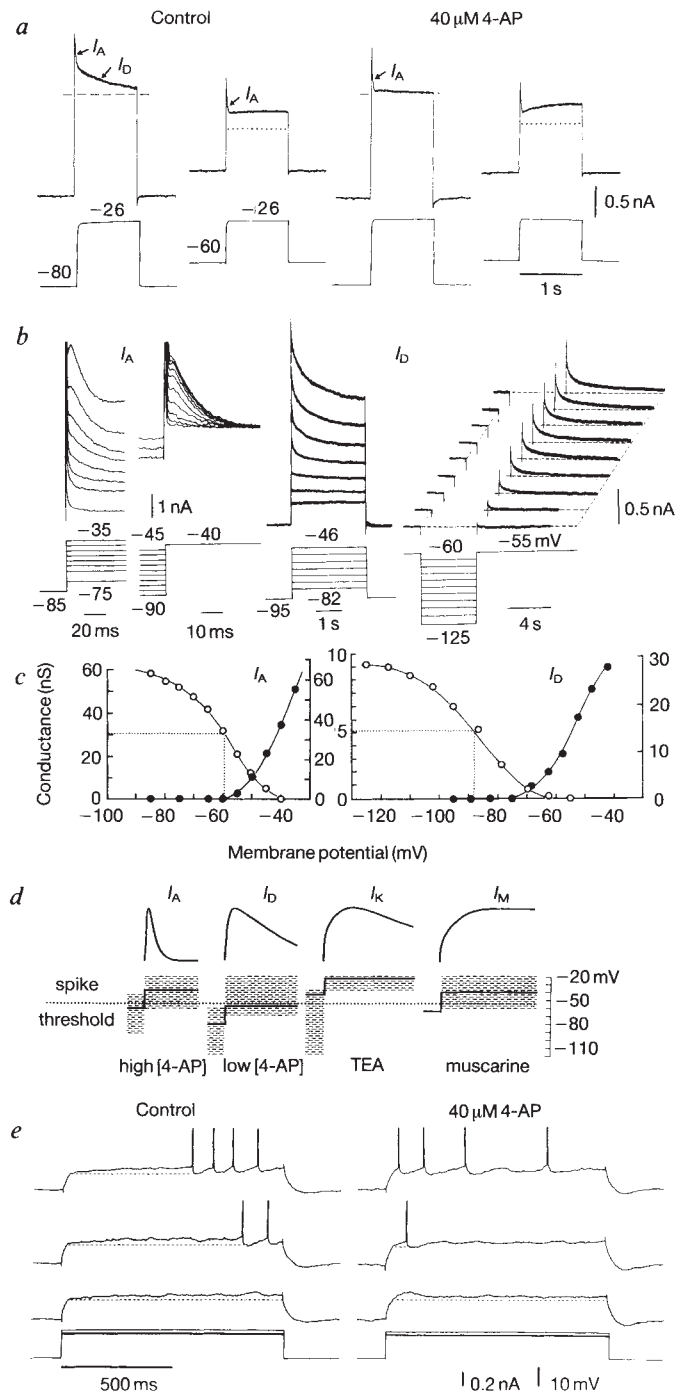


Figure 2*b* shows some of the kinetic properties of I_D compared with those of the more rapidly inactivating A current (I_A) (ref. 10) previously described in dissociated hippocampal cells^{11,12}. The two can be separated because I_A is very much less sensitive to 4-AP, requiring $1-5 \text{ mM}$ for a block. Thus, in the presence of $40 \mu\text{M}$ 4-AP, I_A remains as a brief outward current at the start of depolarization (Fig. 2*a*), inactivating in about 50 ms, as opposed to several seconds for I_D (Fig. 2*b*). Furthermore, the threshold for both activation and inactivation of I_A lie $15-20$ mV positive to those of I_D (Fig. 2*c*).

The A current is apparently too brief to account for the long delay in firing shown in Fig. 1. This is confirmed in Fig. 2*e*. In the presence of a low concentration of 4-AP—sufficient to block I_D but not I_A —the initial delay was drastically shortened. The residual effect of the unblocked A current is then relegated to an initial brief hesitation of about 50 ms, as opposed to the several hundred-ms hesitation induced by I_D . Other K^+ currents

previously described in these cells such as the M current (I_M)¹⁴ and the slow delayed rectifier current (I_K)^{12,13,15} do not contribute to this aspect of firing behaviour: I_K is blocked by TEA and I_M by muscarine, but neither TEA nor muscarine reduced the initial delay. It seems that I_D is the only one of the four voltage-gated K^+ currents¹⁵ with the right properties to generate delayed excitation (Fig. 2*d*). It activates at potentials that are more negative than the spike threshold and inactivates very slowly. Although I_A also activates at a subthreshold level, its inactivation is too fast to account for the long discharge delays. I_K inactivates slowly, like I_D , but its activation range prevents it from contributing to the subthreshold ramp. And, finally, I_M lacks inactivation, although it activates in the subthreshold range.

These conclusions are supported by the observation that the delayed excitation resisted 1 mM Ba^{2+} (which blocks I_M)¹⁴, and 2 mM Cs^+ (which blocks the mixed Na^+/K^+ current I_Q)¹⁴. The

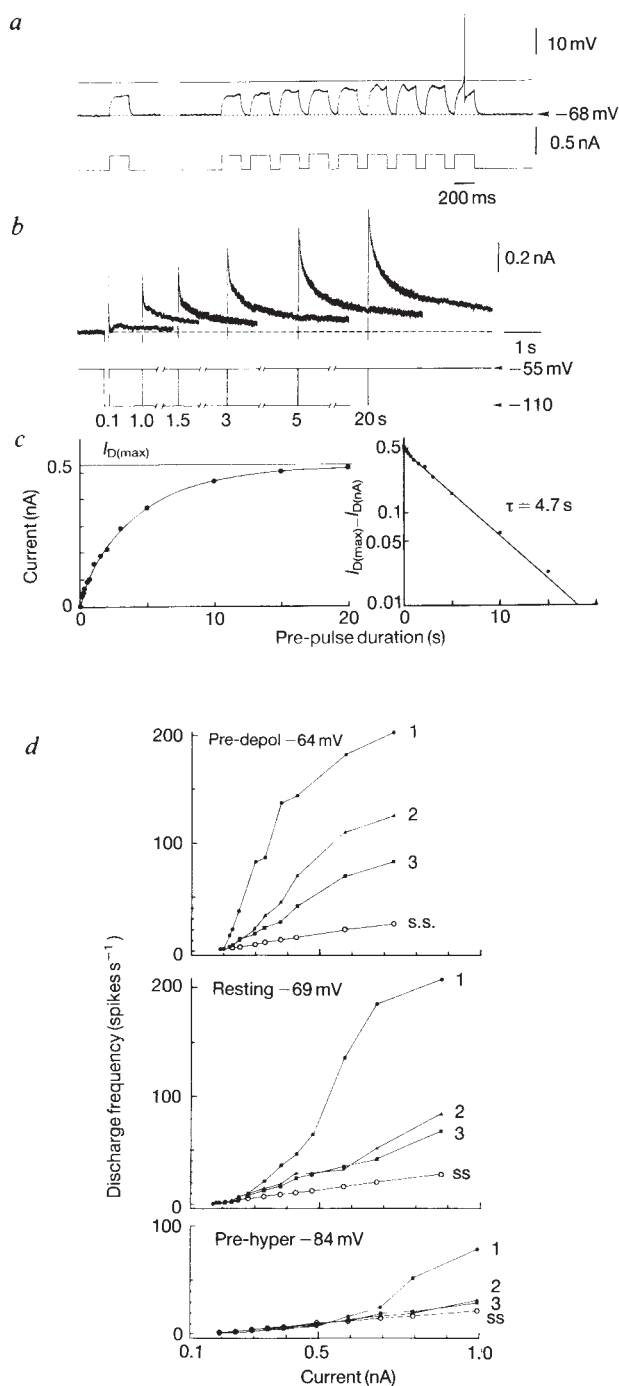


Fig. 3 I_D enables the cell to integrate separate depolarizing episodes, and makes the encoding properties of the cell vary with the resting potential. *a*, A single depolarizing pulse failed to depolarize the cell to the spike threshold (horizontal line), but when a series of identical pulses were injected, the response gradually increased, until the cell fired a spike (truncated). The dotted line marks the resting potential. *b*, Voltage-clamp records, showing that I_D recovers from inactivation very slowly. *c*, Plots of data from *b*. The semi-logarithmic plot (right) shows that I_D recovers with a time constant of about 5 s. *d*, The ability of the cell to transform depolarizing current into discharge was altered by the membrane potential before activation (pre-depol, pre-depolarization; pre-hyper, pre-hyperpolarization). Long (10 s) current pulses were injected, eliciting spikes. The membrane potential before activation was altered by steady current injection. The instantaneous spike frequency for the first three interspike intervals (1, 2, 3) and the average frequency for the steady-state discharge (s.s., 8–10 s after the onset of the pulse) are plotted against the total injected current. **Methods.** *a* and *d*, Normal a.c.s.f. *b* and *c*, Conditions as in Fig. 2*b*.

delay and I_D also resisted 1 mM 9-anthracene carboxylate, which was shown to block the chloride current $I_{Cl(V)}$ (ref. 16) under voltage clamp.

The properties of I_D suggest that it will enable the cell to integrate depolarizing synaptic input over very long times, far exceeding the time span of classical temporal summation. This prediction was tested using direct current injections instead of synaptic input (Fig. 3*a*). When a series of identical pulses were given, the response gradually increased until the cell fired. This behaviour probably depends on I_D , as it was blocked by 30 μ M 4-AP.

The slow inactivation kinetics of I_D enable it to accumulate inactivation over long periods. The recovery from inactivation must also be slow, otherwise I_D would recover in between pulses and there would be no build-up of the response. When measured in voltage-clamp experiments, the recovery of I_D from inactivation proved to be exceedingly slow (time constant about 5 s; Fig. 3*b,c*). In comparison, the time constants of recovery for I_A and I_K were only about 10 ms and 500 ms respectively.

I_D also makes the encoding function of the cell sensitive to the prevailing membrane potential. As I_D is partly inactivated at the normal resting potential, any hyperpolarization or depolarization will change the available I_D and hence the excitability and firing pattern of the cell. In Fig. 3*d*, the discharge frequency of a cell is plotted against the strength of the injected current. At the normal resting potential, the curves for the first three interspike intervals show a characteristic low-gain range for weak currents¹. Hyperpolarization, which enhances I_D , increased the low-gain range several-fold, whereas it disappeared when the cell is depolarized to inactivate I_D . Thus, the input/output relationship of the cell seems to be modulated by I_D .

In conclusion, I_D may have a different role from the classical ones of other K^+ currents, such as the delayed rectifier (spike repolarization), A current (control of maintained repetitive firing⁷), or Ca^{2+} -activated K^+ currents (after-hyperpolarizations and adaptation). In molluscan neurons, a similar delayed excitation seems to be important for the delay of the inking reflex in *Aplysia*^{2,3}, and for slow rhythmic swimming movements in *Tritonia*⁴. In vertebrates, delayed excitation has recently been described in solitary tract neurons that participate in the control of the respiratory rhythms⁵. In the hippocampus, I_D may mediate temporal integration and voltage-dependent modulation of encoding properties. In addition, I_D seems to be fast enough to participate in spike repolarization¹⁷, and presynaptic spike broadening due to block of I_D may mediate the increase in transmitter release which is seen with small doses of 4-AP^{18,19}.

I thank Drs Paul R. Adams and Per Andersen for laboratory space, support and comments. This work was supported by the Norwegian Medical Research Council, a Fogarty International Fellowship and a NIH grant to Dr Adams.

Received 5 July; accepted 18 October 1988.

1. Lanthorn, T., Storm, J. & Andersen, P. *Expl Brain Res.* **53**, 431–443 (1984).
2. Byrne, J. H., Shapiro, E., Dieringer, N. & Koester, J. *J. Neurophysiol.* **42**, 1233–1250 (1979).
3. Byrne, J. H. *J. Neurophysiol.* **43**, 651–668 (1980).
4. Getting, P. A. *J. Neurophysiol.* **49**, 1036–1050 (1983).
5. Dekin, M. S., Getting, P. A. & Johnson, S. M. *J. Neurophysiol.* **58**, 195–214 (1987).
6. Yarom, Y. & Llinas, R. *J. Neurophysiol.* **7**, 1166–1177 (1987).
7. Rudy, B. *Neuroscience* **25**, 729–749 (1988).
8. Dubois, J. M. *Prog. Biophys. molec. Biol.* **42**, 1–20 (1983).
9. Stansfield, C., Marsh, S. J., Halliwell, J. V. & Brown, D. A. *Neurosci. Lett.* **64**, 299–304 (1986).
10. Connor, J. A. & Stevens, C. F. *J. Physiol. Lond.* **213**, 31–53 (1971).
11. Segal, M., Rogawsky, M. A. & Barker, J. L. *J. Neurosci.* **4**, 604–609 (1984).
12. Segal, M. & Barker, J. L. *J. Neurophysiol.* **51**, 1409–1433 (1984).
13. Numann, R. E., Wadman, W. J. & Wong, R. K. S. *J. Physiol. Lond.* **393**, 331–353 (1987).
14. Halliwell, J. V. & Adams, P. R. *Brain Res.* **250**, 71–92 (1982).
15. Storm, J. F. *Biophys. J.* **53**, 148a (1988).
16. Madison, D. V., Malenka, R. C. & Nicoll, R. A. *Nature* **321**, 695–697 (1986).
17. Storm, J. F. *Eur. J. Neurosci. Suppl.* **1**, 42 (1988).
18. Thesleff, S. *Neuroscience* **5**, 1413–1419 (1980).
19. Buckle, P. J. & Haas, H. L. *J. Physiol., Lond.* **326**, 109–122.
20. Mayer, M. & Sugiyama, K. *J. Physiol., Lond.* **396**, 417–433 (1987).

Noradrenaline contracts arteries by activating voltage-dependent calcium channels

Mark T. Nelson, Nicholas B. Standen*,
Joseph E. Brayden & Jennings F. Worley III†

Department of Pharmacology,
University of Vermont College of Medicine, Burlington,
Vermont 05405, USA

* Department of Physiology, University of Leicester,
Leicester LE1 7RH, UK

† Department of Pharmacology and Toxicology,
West Virginia University, Morgantown, West Virginia, USA

Noradrenaline (NA) regulates arterial smooth muscle tone and hence blood vessel diameter and blood flow¹. NA apparently increases tone by causing a calcium influx through the cell membrane¹⁻³. Two calcium influx pathways have been proposed: voltage-activated calcium channels and NA-activated calcium-permeable channels that are voltage-insensitive¹. Although voltage-activated calcium channels have been identified in arterial smooth muscle^{4,5}, voltage-insensitive calcium channels activated by NA have not. We show here that NA contractions of rabbit mesenteric arteries increase with depolarization. The increase parallels the elevation of open-state probability (P_o) of single, voltage-dependent calcium channels. The action of noradrenaline can be explained by NA-activating voltage-dependent calcium channels, rather than by opening a second type of channel. We show directly that NA increases the open-state probability of single calcium channels. Thus, in the presence of NA, calcium entry through voltage-dependent calcium channels can regulate smooth muscle tone at physiological membrane potentials. These results may have relevance to pathophysiological conditions such as hypertension.

Maintained tone of arterial smooth muscle in response to high external potassium (K_o^+) or noradrenaline depends on calcium influx^{1,2}. If NA activates a voltage-insensitive calcium-permeable channel, then force development in response to applied NA should decrease with depolarization as the driving force for calcium entry is reduced. But we find that depolarization by K_o^+ markedly increases contractions of mesenteric arterial rings (Fig. 1a). In 40 arteries, a 20 mV depolarization (K_o^+ raised from 5 to 20 mM) increased NA contractions 6.62-fold (± 0.85 , s.e.m.). In the absence of NA, force develops at potentials positive to -45 mV (20 mM K_o^+) (Fig. 1a)^{6,7}. The calcium channel opener Bay R 5417 (ref. 8) also increased NA contractions and shifted the threshold for depolarization-induced contraction to more negative potentials (Fig. 1b). The calcium-channel inhibitors nisoldipine and cadmium⁹ reduced both high- K_o^+ and NA contractions with the same apparent inhibition constant (0.95 nM at -45 mV and 50 μ M, respectively).

Our data indicate that maintained contractions to NA depend on calcium influx through voltage-dependent calcium channels. To gain insights into the relationship between the voltage-dependence of calcium channels and of force, the effect of membrane potential (V_m) on both single-channel current (i) and on channel P_o must be known; this can be most accurately determined at the single-channel level. Calcium influx will be proportional to NiP_o , where N is the total number of calcium channels. Depolarization reduced the unitary current through single channels (Fig. 2), the slope conductance being larger at potentials close to the resting potential (37.0 pS at -60 mV) than at 0 mV (18.5 pS) (Fig. 2d).

P_o for single calcium channels increased with depolarization. The relation between V_m and P_o can be described by

$$P_o = P_{\max} / (1 + \exp [(V_{0.5} - V_m) / k]) \quad (1)$$

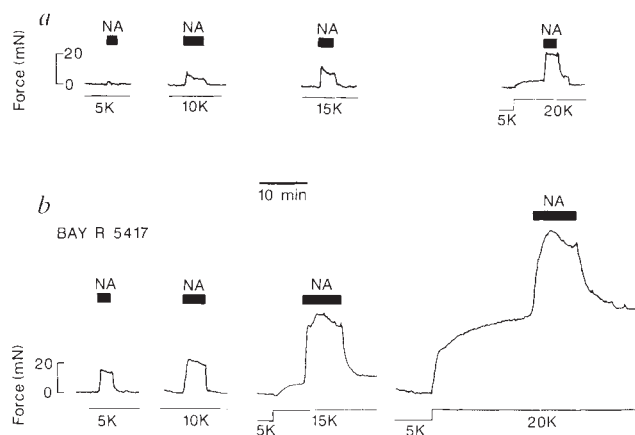


Fig. 1 Membrane depolarization and Bay R 5417 increase noradrenaline contractions. *a*, Effect of depolarization on noradrenaline contractions. Membrane potential (V_m) was depolarized by elevating K_o^+ from 5 mM to 10 mM, 15 mM, 20 mM. Noradrenaline (2.5 μ M) was applied at times indicated by the solid bars. To prevent complications from β -adrenergic stimulation and neuronal uptake of NA, propranolol (1 μ M) and desipramine (0.3 μ M) were included in the bathing solution in this experiment. *b*, Effects of Bay R 5417 on voltage-dependent contractions with and without noradrenaline. Bay R 5417 (100 nM) was included in all solutions. Nisoldipine (500 nM) reduced the contraction in 20 mM K_o^+ , NA, Bay R 5417 by more than 92%.

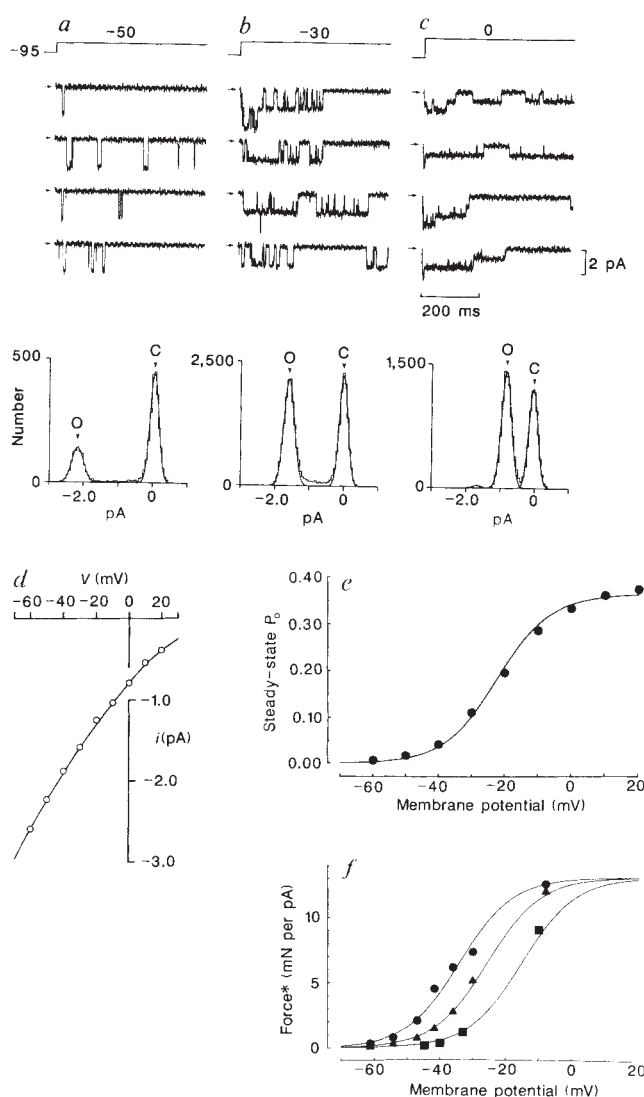
Methods. Male rabbits were anaesthetized by pentobarbital and then exsanguinated. Mesenteric arteries (200–400 μ m outer diameter) were rapidly removed and placed in (mM): NaCl 137, KCl 4.56, KH_2PO_4 0.44, NaH_2PO_4 0.42, $NaHCO_3$ 4.17, glucose 11.11, EGTA 0.05, $MgCl_2$ 1, $CaCl_2$ 1.8, HEPES/NaOH 10, pH 7.4 (37 °C). Isometric contractions of ring segments of arteries were measured using a resistance vessel myograph⁵. To measure membrane potential, smooth muscle cells in the intact artery were impaled with a microelectrode under isometric conditions. Microelectrodes were filled with 0.5 M KCl and had tip resistances of 100–150 Mohms. The membrane potentials in different potassium solutions were: 5 mM K^+ , -64.8 mV; 10 mM K^+ , -58.7 mV; 15 mM K^+ , -50.0 mV; 20 mM K^+ , -44.7 mV; 25 mM K^+ , -40.0 mV; 30 mM K^+ , -34.5 mV (means of 5–10 determinations). In 0.5 μ M NA, V_m in 5, 10, 15, 20, 25 and 30 mM K_o^+ were -61.2 mV, -54.3 mV, -47.0 mV, -41.7 mV, -36 mV and -31 mV, respectively; that is, NA depolarized the arteries by about 3–4 mV ($n=5-10$). Similar relationships between external K^+ and V_m have been reported previously^{6,7}.

where $P_{\max}$ is the maximal P_o at positive potentials, $V_{0.5}$ is the potential where P_o is one-half $P_{\max}$, and k is the steepness factor, an indicator of the voltage-sensitivity of calcium channels. The solid curve in Fig. 2e has $V_{0.5} = -22.5$ mV and $k = 8.0$ mV. In seven experiments, k and $V_{0.5}$ ranged from 7.0 to 8.0 mV, and from -15 to -23 mV respectively.

To compare the voltage-dependence of steady-state P_o (Fig. 2e) and force, the effect of depolarization on calcium influx through the open channel must be taken into account (Fig. 2d). To do this, force is divided by the mean unitary current at each V_m and is then referred to as force*. The relationships between V_m and force* in control, NA, and NA + Bay R 5417 are shown in Fig. 2f and can be empirically fitted by equation (1) with $k = 8.0$ mV. The similarity between V_m - P_o (Fig. 2e) and V_m -force* relationships suggests that the elevation in force with depolarization reflects the increase in P_o of voltage-activated calcium channels.

We have also demonstrated the activation by NA of single voltage-dependent calcium channels (Fig. 3). Single calcium channels were recorded from a cell-attached patch before and after the addition of NA (10 μ M) to the bath solution. NA caused a dramatic increase in P_o (4 experiments), whereas the mean unitary current (-1.18 pA) was unchanged. As NA was

Fig. 2 Effect of membrane potential on single calcium channels and force. *a*, On-cell single calcium channel recordings and amplitude histograms at -50 mV. The membrane potential was stepped from a holding potential of -95 mV to -50 mV. The top record shows the voltage step and the four records beneath show examples of unitary current recordings. The arrows to the left of each record indicate the current level when the channels were closed. The unitary current was measured by fitting gaussian curves to the amplitude histogram using a least-squares algorithm and was -2.22 pA at -50 mV. C and O correspond to the closed and first open levels respectively. The amplitude of the unitary current measured at the resting potential of the cell (-1.24 pA) changed by less than 3% during the course of the experiment, suggesting that the resting potential remained stable. The resting potential of the cell was estimated from the change in unitary current after excision of the patch using the single-channel current-voltage relation. In the excised patch at 0 mV the unitary current was -0.77 pA, giving a value for the resting potential of -20 mV. All values of membrane potential are given as true membrane potential, namely applied potential plus resting potential. Barium was used as the charge carrier to maximize the value of the unitary current^{21,22}. *b*, Single channel recordings and histogram for a voltage step to -30 mV. Unitary current was -1.58 pA. *c*, Single channel recordings and histogram for a voltage step to 0 mV. Unitary current was -0.79 pA. *d*, Single channel current-voltage relationship. Mean unitary currents were determined from amplitude histograms. Single smooth muscle cells were enzymatically isolated from rabbit mesenteric artery as previously described by Worley *et al.*⁴. The external (pipette) solution was (mM): 80 BaCl₂, 10 HEPES/BaOH, pH 7.4 and the bath solution was (mM): 120 NaCl or sodium aspartate, 20 CsCl, 10 HEPES/NaOH, 5 EGTA, pH 7.0. To have enough channel openings to analyse statistically, either Bay K 8644 ($1 \mu\text{M}$) or its isomer Bay R 5417 ($0.25 \mu\text{M}$) was included in the bathing solution. Current and voltage were recorded on a Dagan or Axopatch patch clamp amplifier and stored on video tape. The tape was later replayed through an 8-pole Bessel filter (-3 dB at 500 or 1,000 Hz), digitized at 2–4 KHz, and analysed by computer. Capacity and leakage currents associated with changes in potential were removed, partly by analog means and partly digitally, by averaging records free of channel openings and subtracting this average from all records. *e*, Relationship between membrane potential and steady-state open-state probability, P_o , measured over the duration of 1-second depolarizing pulses as $\sum_{j=1}^N(t_j)/TN$, where t_j s are the total times spent at the current levels corresponding to $j = 0, 1 \dots N$ channels open, T is the total time, and N is the maximum number of channels seen in the patch under conditions where P_o was high. P_o is plotted as the closed circles and was fitted using equation (1), with the maximal P_o , the midpoint of the activation curve and the steepness factor being 0.36, -22.5 mV and 8.0 mV, respectively (solid curve). We regard the shape and steepness of the voltage-dependence of P_o (given by k) as much more significant than the absolute position of the relation on the voltage axis, because we would expect this to be shifted to more positive voltages as a result of the high divalent concentration in the pipette and to be shifted to more negative potentials by Bay K 8644 or Bay R 5417. *f*, Relationship between membrane potential and force* in control (squares), NA ($0.5 \mu\text{M}$, triangles), and NA+Bay R 5417 (100 nM, circles). The lines are drawn to equation (1) using the same steepness factor (8.0 mV). The maximal force* (13.1 mN per pA) used for the fits represents the maximal force measured in this experiment. The midpoints of these relationships were: in control, -15 mV; in NA, -25.5 mV; in NA+Bay R 5417, -34 mV. Force* is the measured force divided by the mean unitary current (from *d*) at that V_m , assuming the slope of the current-voltage relation in physiological $[\text{Ca}^{2+}]$ is essentially similar to that in 80 mM barium, which is reasonable as the divalent cation gradient is very steep in both cases²². Arteries were depolarized by 10, 20, 25, 30 and 85 mM K_o⁺. Force values at each K_o⁺ level are the means of four determinations performed on two arterial segments. The relationship between V_m and force* in Bay R 5417 without NA can be fitted with a midpoint of -27 mV and a steepness factor of 8.0 mV (not shown). As with the V_m - P_o relationship, we attach significance to the voltage sensitivity of force, and not to the maximal force or the midpoints, which depend on a number of factors.



applied to the bath solution and not directly to the membrane patch, these results suggest the involvement of an intracellular second messenger. This result is complemented by the recent finding that NA increases whole cell calcium currents of smooth muscle¹⁰.

Our results indicate that any agent that modulates V_m should affect tone through voltage-dependent calcium channels, for example, potassium-channel openers, fluctuations in K_o⁺ (Fig. 1) or inhibitors of the (Na⁺ + K⁺)ATPase. We tested the effects of a reversible (Na⁺ + K⁺)ATPase inhibitor, dihydro-ouabain (DHO), on contraction and V_m . In the absence of NA, DHO depolarized arteries by 11.4 ± 1.8 mV ($n = 5$, s.e.m.) and, as expected (Fig. 1), this degree of depolarization did not affect force (see also refs 6 and 7). In the presence of NA ($0.5 \mu\text{M}$), however, DHO depolarized and increased force (Fig. 4). After

removal of NA, but in the continued presence of DHO, the artery relaxed completely whereas the membrane hyperpolarized only slightly (-49.3 to -53.7 mV), indicating that any change in intracellular sodium caused by DHO did not cause contraction in the absence of NA. In this experiment the voltage dependence of force (Fig. 4b) was very similar to that of calcium channels (Fig. 2).

The role of calcium channels in DHO-induced contractions is confirmed by the effects of Bay R 5417 and nisoldipine (Fig. 4c). In NA, DHO increased force 2.8-fold (upper trace). Bay R 5417 had only a slight effect on force under resting conditions and, as expected, increased the NA contraction (lower trace). Bay R 5417 increased the contraction resulting from (Na⁺ + K⁺)ATPase inhibition by DHO (lower trace). Force development was abolished by nisoldipine, indicating that all of the

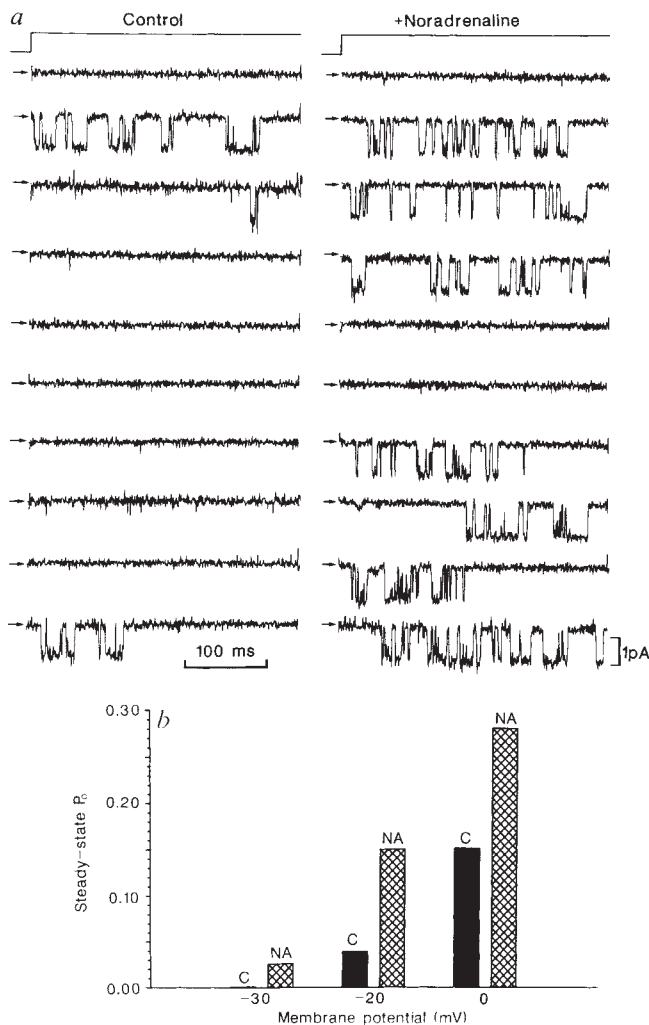


Fig. 3 Noradrenaline opens single, voltage-dependent calcium channels. *a*, On-cell single calcium channel recordings before and after addition of noradrenaline ($10 \mu\text{M}$) to the bathing solution. The membrane potential was stepped from a holding potential of -95 mV to -20 mV . Membrane potential is the pipette potential plus the cell's membrane potential (-20 mV). The cell's membrane potential was estimated from the single channel current-voltage relationship (see Fig. 2). NA did not affect the unitary current, indicating that it did not affect the membrane potential. This is not surprising because the experimental solutions do not contain potassium or calcium. Records were filtered at 1 kHz . This patch contained one channel, as judged by a lack of two simultaneous openings when P_o was high ($+20 \text{ mV}$ in NA). Bay R 5417, 250 nM . *b*, Noradrenaline and membrane depolarization increase steady-state open-state probability. Steady-state P_o was measured as described in Fig. 2e (64–152 test pulses were averaged under each condition) and was determined before and after the application of NA to the extrapipette solution.

tone was dependent on calcium entry through voltage-dependent calcium channels.

These results indicate that NA contracts arterial smooth muscle by opening voltage-dependent calcium channels. We propose that voltage-activated calcium channels mediate vasoconstriction in response to a variety of substances¹¹ such as serotonin¹² and endothelin¹³. The nature of the transduction signal from alpha-receptors to voltage-dependent calcium channels is not known, but could involve inositol trisphosphate or diacylglycerol formation^{14–18}.

Our findings suggest mechanisms that could increase peripheral resistance and so contribute to hypertension. We would argue that any condition that increases P_o of voltage-

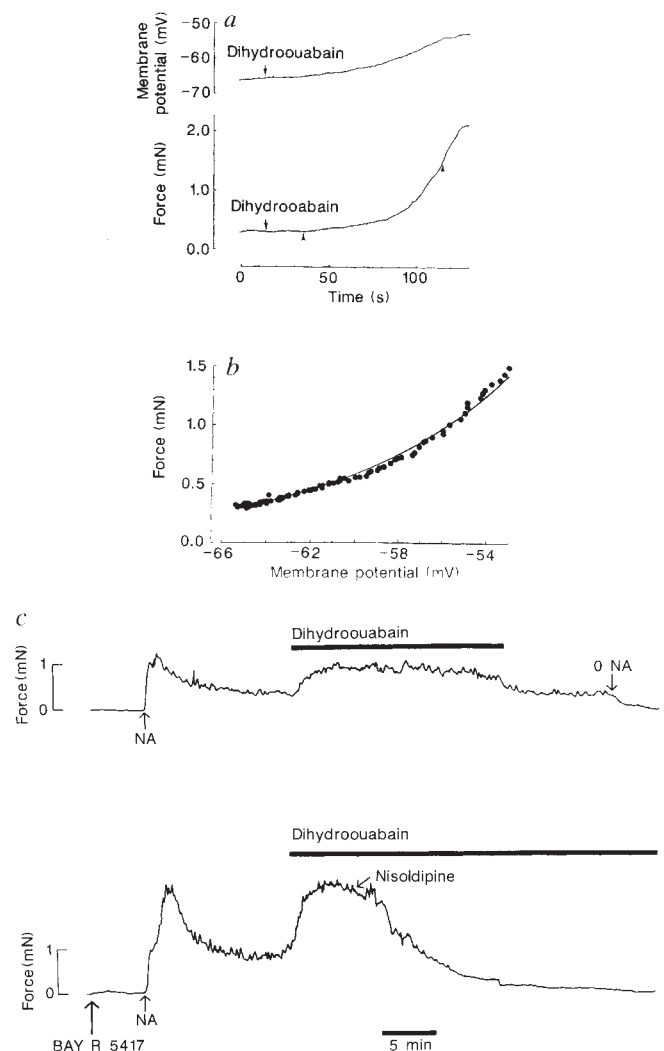


Fig. 4 $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ inhibition contracts arteries by way of voltage-dependent calcium channels. *a*, $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ inhibition increases force in the presence of NA. Original records showing the effect of dihydro-ouabain ($3 \mu\text{M}$) on membrane potential and force of an artery bathed in NA ($0.5 \mu\text{M}$). NA by itself induced a contraction of about 0.32 mN and depolarized the artery from -68.1 mV to -66.2 mV . *b*, Relationship between membrane potential and force derived from the data between the arrows shown in (*a*). Values of force and V_m were taken from the digitized data averaged over 1 s intervals. The relationship was fitted by the equation $B \exp(V_m/k)$ (solid line), where k , the steepness factor, was 8.0 mV (range 7.5 – 9.0 mV , 5 experiments) and B , the scaling factor, was $1,242 \text{ mN}$. *c*, Contraction resulting from $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ inhibition depends on calcium influx through voltage-activated calcium channels. Tracings are of original force measurements. The upper and lower traces are from the same experiment. NA, $1.5 \mu\text{M}$; Bay R 5417, 125 nM ; DHO, $3 \mu\text{M}$; nisoldipine, $1 \mu\text{M}$. Membrane potential and force were recorded simultaneously in preparations mounted in a myograph. Mesenteric arteries were superfused at a flow rate of 2 ml min^{-1} and NA, DHO, Bay R 5417 and nisoldipine were applied through the superfusate solution.

activated calcium channels *in vivo*, for example elevated NA (ref. 19) or depolarization, would increase the tone of arterial smooth muscle. Depolarization might occur as a result of decreased $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ activity (compare Fig. 4), shown to be associated with essential hypertension in some cases²⁰.

We thank M. Martin, D. Van Gerbig, L. Hazlehurst for technical assistance, Drs R. Gibbons and J. Patlak for discussion and Dr A. Scriabine (Miles Institute of Preclinical Phar-

macology) for nisoldipine, Bay K 8644, and Bay R 5417. This work was supported by grants from the National Science Foundation (M.T.N.), American Heart Association (M.T.N.), Miles Laboratories (M.T.N.), a Burroughs Wellcome Travel grant (M.T.N.), Wellcome and Nuffield Travel grants (N.B.S.), Medical Research Council (N.B.S.), and the NIH (J.E.B.). J.F.W. was supported by a Muscular Dystrophy fellowship and the Pharmaceutical Manufacturers Association Foundation. N.B.S. is supported by a Wellcome Trust Research Leave fellowship and M.T.N. is an established investigator of the American Heart Association.

Received 30 June; accepted 27 October 1988.

1. Bulbring, E. & Tomita, T. *Pharmacol. Rev.* **39**, 49-96 (1987).
2. Bolton, T. B. *Physiol. Rev.* **59**, 606-718 (1979).
3. Janis, R. A., Silver, S. C. & Triggle, D. J. *Adv. Drug Res.* **16**, 311-536 (1987).
4. Worley III, J. F., Deitmer, J. & Nelson, M. T. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5746-5750 (1986).
5. Benham, C. D., Hess, P. & Tsien, R. W. *Circulation Res.* **61**, 10-16 (1987).
6. Bolton, T. B., Lang, R. J. & Takewaki, T. *J. Physiol., Lond.* **351**, 549-572 (1984).
7. Mulvany, M. J., Nilsson, H. & Flatman, J. A. *J. Physiol., Lond.* **332**, 363-373 (1982).
8. Franckowiak, G., Bechem, M., Schramm, M. & Thomas, G. *Eur. J. Pharmac.* **114**, 223-226 (1985).
9. Standen, N. B., Worley, J. F. & Nelson, M. T. *Biophys. J.* **53**, 231a (1988).
10. Benham, C. D. & Tsien, R. W. *J. Physiol., Lond.* **404**, 767-784 (1988).
11. Fleming, W. W. A. *Rev. Pharmac. Tox.* **20**, 129-149 (1980).
12. Neild, T. O. *J. Physiol., Lond.* **386**, 19-30 (1987).
13. Yanagisawa, M. *et al. Nature* **332**, 411-415 (1988).
14. Han, C., Abel, P. W. & Minneman, K. P. *Nature* **329**, 333-335 (1987).
15. Hashimoto, T., Hirata, M., Itoh, T., Kanmura, Y. & Kuriyama, H. *J. Physiol., Lond.* **370**, 605-618 (1986).
16. Somlyo, A. V., Bond, M., Somlyo, A. P. & Scarpa, A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5231-5235 (1985).
17. Fish, D., Sperti, G., Colucci, W. & Clapham, D. *Circulation Res.* **62**, 1049-1054 (1988).
18. Nastainczyk, W. *et al. Eur. J. Biochem.* **169**, 137-142 (1987).
19. Bohm, R. O., van Baak, M. A., van Hooff, M. E., Mooy, J. & Rahn, K. H. *J. Hypertension* **5**, 655-662 (1987).
20. Hamlyn, J. M. *et al. Nature* **300**, 650-652 (1982).
21. Nelson, M. T., French, R. J. & Krueger, B. K. *Nature* **308**, 77-80 (1984).
22. Hess, P., Lansman, J. B. & Tsien, R. W. *J. gen. Physiol.* **88**, 293-320 (1986).

Endothelium-derived relaxing factor release on activation of NMDA receptors suggests role as intercellular messenger in the brain

John Garthwaite, Sarah L. Charles
& Russell Chess-Williams*

Departments of Physiology and *Pharmacology,
University of Liverpool, Brownlow Hill,
PO Box 147, Liverpool L69 3BX, UK

In the vascular system, endothelium-derived relaxing factor (EDRF) is the name of the local hormone released from endothelial cells in response to vasodilators such as acetylcholine, bradykinin and histamine. It diffuses into underlying smooth muscle where it causes relaxation by activating guanylate cyclase, so producing a rise in cyclic GMP levels¹⁻³. It has been known for many years that in the central nervous system (CNS) the excitatory neurotransmitter glutamate can elicit large increases in cGMP levels, particularly in the cerebellum where the turnover rate of cGMP is low⁴. Recent evidence indicates that cell-cell interactions are involved in this response⁵. We report here that by acting on NMDA (N-methyl-D-aspartate) receptors on cerebellar cells, glutamate induces the release of a diffusible messenger with strikingly similar properties to EDRF. This messenger is released in a Ca²⁺-dependent manner and its activity accounts for the cGMP responses that take place following NMDA receptor activation. In the CNS, EDRF may link activation of postsynaptic NMDA receptors to functional modifications in neighbouring presynaptic terminals and glial cells.

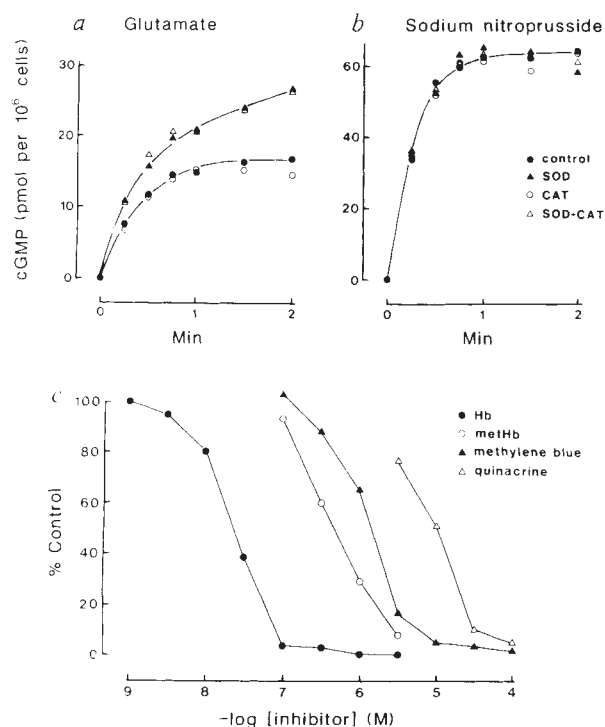


Fig. 1 *a* and *b*, Time courses of cGMP accumulation in rat cerebellar cell suspensions in response to *a*, L-glutamate (100 μM) and *b*, sodium nitroprusside (300 μM) in the absence (●) or presence of superoxide dismutase (30 U ml⁻¹, ▲), catalase (200 U ml⁻¹, ○) or both enzymes together (△). The enzymes were added to the cells 1 min before the effectors. Each point is the mean of two determinations, each being made on different aliquots (1.7 × 10⁶ cells) of a cell suspension prepared from the cerebella of 20 rats. Similar results were obtained in two other experiments. The mean increase (±s.e.m.) in the amplitude of the response to glutamate in the presence of superoxide dismutase after 1 min exposure was 169 ± 18%. *c*, Inhibition of cGMP responses to NMDA (100 μM) by haemoglobin (●), methaemoglobin (○), methylene blue (▲) and quinacrine (△). The cells were exposed to NMDA for 1 min, with the antagonists being added 1 min beforehand. Each point represents the mean of two determinations, each done on different aliquots of 1.5–2 × 10⁶ cells; the data have been expressed as % response in the absence of antagonist. Haemoglobin and methaemoglobin were prepared as described¹¹. **Methods.** These have been described previously^{5,6}. Briefly, 8–9-day-old rat cerebella, depleted of non-responsive dividing cells, were dissociated into single-cell suspensions and incubated at 37 °C in an air-equilibrated medium containing (mM): NaCl, 130; KCl, 5; CaCl₂, 2; MgSO₄, 1.2; Na₂HPO₄, 1.2; Tris-HCl, 15; and glucose, 11, supplemented with 1% bovine serum albumin (essentially fatty acid-free). Cell concentrations (15–20 × 10⁶ cells ml⁻¹) were determined using a haemocytometer. The cells were inactivated in boiling hypotonic buffer and cGMP levels measured by radioimmunoassay.

The experiments used cell suspensions prepared from the cerebellum of 8–9-day-old rats^{5,6}. At all time points the cGMP response to glutamate (100 μM) was increased in the presence of superoxide dismutase (Fig. 1*a*), an enzyme that catalyses the reaction: 2O₂⁻ + 2H⁺ → O₂ + H₂O₂. This effect is due to removal of superoxide ions (O₂⁻) as it was not reproduced or affected by catalase (2H₂O₂ → O₂ + 2H₂O). EDRF is an unstable compound whose half life is prolonged by superoxide dismutase^{7,8}. It has recently been identified as nitric oxide (NO)^{9,10}, which reacts with O₂⁻ to give NO₃⁻ (ref. 9). Sodium nitroprusside and other nitrovasodilators produce NO spontaneously and so mimic EDRF³. In the cerebellar cells, sodium nitroprusside (0.3 mM) elicited a large rise in cGMP (Fig. 1*b*) that was unaffected by superoxide dismutase, catalase or a combination of the two. (At

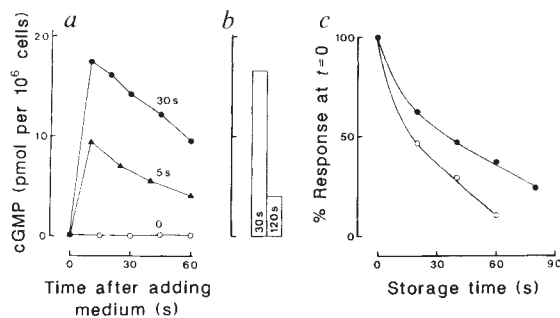


Fig. 2 The suspending medium from NMDA-stimulated cells contains an unstable factor which elicits cGMP accumulation. *a*, NMDA ($100 \mu\text{M}$) was added to cells (incubated with 30 U ml^{-1} superoxide dismutase) and after either 5 s ($\blacktriangle$) or 30 s ($\bullet$) the cells were removed by rapid filtration ($0.45\text{-}\mu\text{m}$ pore Millipore filter) and 0.5 ml medium added immediately ($t=0$) to 0.5 ml 'detector' cells containing superoxide dismutase and sufficient APV ($200 \mu\text{M}$) to prevent the effects of the added NMDA⁶. Portions of the diluted cells were taken for cGMP analysis at the intervals indicated. For the control ($\circ$), the medium from unstimulated cells was used. In a similar experiment (*b*), the initial exposure periods to NMDA were 30 s and 120 s; only peak responses (in both cases 10 s after addition of the medium, the earliest time point examined) are shown. In (*c*), the medium was taken (as above) from cell suspensions exposed to NMDA ($100 \mu\text{M}$) for 30 s in the absence ($\circ$) or presence ($\bullet$) of superoxide dismutase (30 U ml^{-1}) and stored at 37°C ; aliquots of the medium were then added to equal volumes of detector cells (containing superoxide dismutase and $200 \mu\text{M}$ APV) at various intervals. Each point represents cGMP levels in the detector cells 10 s after adding the medium, expressed as % amplitude of the levels evoked by the first aliquots (11 and 9 pmol cGMP per 10^6 cells in the presence and absence of superoxide dismutase respectively). The data are plotted against time after adding the first aliquot, which corresponded to 45 s after the initial addition of NMDA. Note that, in the presence of superoxide dismutase, the 75% loss of activity in the medium after 80 s storage agrees well with the 75% loss of activity of the medium observed when the exposure of the cells to NMDA was prolonged by 90 s (*b*); this would suggest that the cells themselves do not contribute significantly to the inactivation of the factor under these conditions. Samples of all the media were checked for the presence of cGMP but none was detected, consistent with previous evidence that cGMP itself is not released in significant amounts⁵.

relatively high sodium nitroprusside concentrations, the decay of NO would not be expected to be limiting). Together with evidence that sodium nitroprusside and glutamate elevate cGMP through the same pool of guanylate cyclase (GC)⁵, this indicates that the increased response to glutamate in the presence of superoxide dismutase is unlikely to be due to the removal of a deleterious effect of O_2^- on guanylate cyclase, or on the cells containing it.

The increase in cGMP evoked by glutamate, in a near-maximally effective concentration ($30 \mu\text{M}$), seems to be mediated exclusively through NMDA receptors in these cells⁶. The cGMP response to NMDA ($100 \mu\text{M}$) was accordingly also elevated in the presence of superoxide dismutase: after 1-minute exposure, the mean increase ($\pm\text{s.e.m.}$) was $168 \pm 8\%$ ($n=12$), a value not significantly different from the corresponding increase in the response to glutamate (see Fig. 1 legend). As it is a more selective agonist whose actions are not complicated by cellular uptake⁶, NMDA was the main effector used in subsequent experiments.

EDRF is potentially inhibited by haemoglobin and less so by methaemoglobin⁹⁻¹¹. The inhibition is explained by the ability of haemoglobin and methaemoglobin to bind NO. In the cerebellar cells (Fig. 1c), the cGMP response to $100 \mu\text{M}$ NMDA was blocked by haemoglobin (50% inhibitory concentration, $\text{IC}_{50} = 25 \text{ nM}$) and, more weakly, by methaemoglobin ($\text{IC}_{50} = 400 \text{ nM}$). Methylene blue, which prevents activation of guanylate cyclase

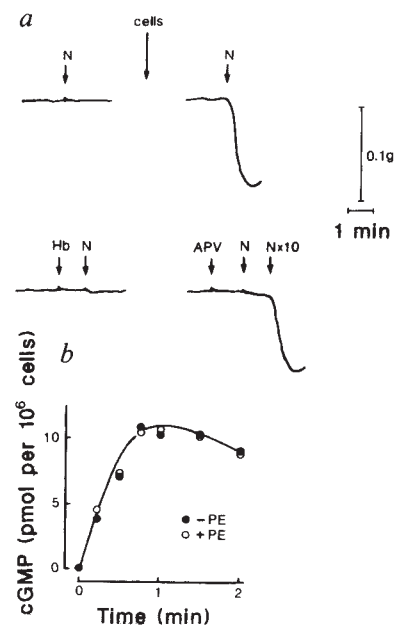


Fig. 3 *a*, NMDA causes smooth muscle relaxation in the presence of cerebellar cells. An open ring of rat aorta was pre-contracted with 100 nM phenylephrine which was present throughout the rest of the experiment. Addition of NMDA (N, $100 \mu\text{M}$) failed to elicit a change in tension. The bathing medium was then changed to one in which cerebellar cells were suspended at a concentration of $19 \times 10^6 \text{ cells ml}^{-1}$. Application of NMDA then caused a prompt relaxation corresponding in magnitude to 45% of the phenylephrine-induced contracture. The muscle was washed twice and cerebellar cells readmitted. Addition of haemoglobin (Hb) (100 nM) had no effect, but abolished the NMDA-induced relaxation. After washing and addition of fresh cells, APV (D-isomer, $50 \mu\text{M}$) inhibited the response to $100 \mu\text{M}$ NMDA in a manner which could be overcome by increasing the NMDA concentration to 1 mM . *b*, The cGMP response to NMDA in cerebellar cell suspensions is unaffected by the contracting agent. Cells were exposed to NMDA ($100 \mu\text{M}$) in the absence ($\bullet$) or presence ($\circ$) of 100 nM phenylephrine (added 1 min beforehand). Each point is the mean of two determinations, each done on different aliquots of 2×10^6 cells.

Methods. A 5 mm segment of rat thoracic aorta was cut longitudinally to form a sheet. The endothelium was removed by blotting on filter paper. The muscle was set up under 1 g tension in 0.5 ml the same medium used to incubate the cerebellar cells (see Fig. 1) and continuously stirred with a magnetic flea at 37°C . The preparation was allowed to equilibrate for 1–2 h (with several changes of bath fluid) before being contracted submaximally with phenylephrine. Isometric tension was recorded. Drugs were added in 2.5 or $5 \mu\text{l}$ volumes.

by EDRF^{11,12}, was also effective ($\text{IC}_{50} = 1.4 \mu\text{M}$), as was quina-crine ($\text{IC}_{50} = 10 \mu\text{M}$), one of several non-specific EDRF antagonists². Haemoglobin ($1 \mu\text{M}$) and methylene blue ($10 \mu\text{M}$) also inhibited the cGMP response to $100 \mu\text{M}$ glutamate by $>95\%$ and $>85\%$ respectively (two experiments).

To determine whether an EDRF-like factor is released, cells (incubated with superoxide dismutase) were stimulated with $100 \mu\text{M}$ NMDA and after 30 seconds the suspending medium was removed by rapid filtration and added to 'detector' cells containing superoxide dismutase and the NMDA receptor blocker, D,L-2-amino-5-phosphonovalerate (APV). There was a prompt but transient rise in cGMP in the detector cells (Fig. 2a), confirming that the medium contained a factor capable of stimulating cGMP production. The medium from cells not exposed to NMDA (Fig. 2a) or from cells exposed to NMDA in the presence of APV (see Fig. 4d) was ineffective.

The use of differing periods of exposure of the 'donor' cells to NMDA indicated that the factor initially accumulates in the medium (Fig. 2a) and then its activity subsides (Fig. 2b), imply-

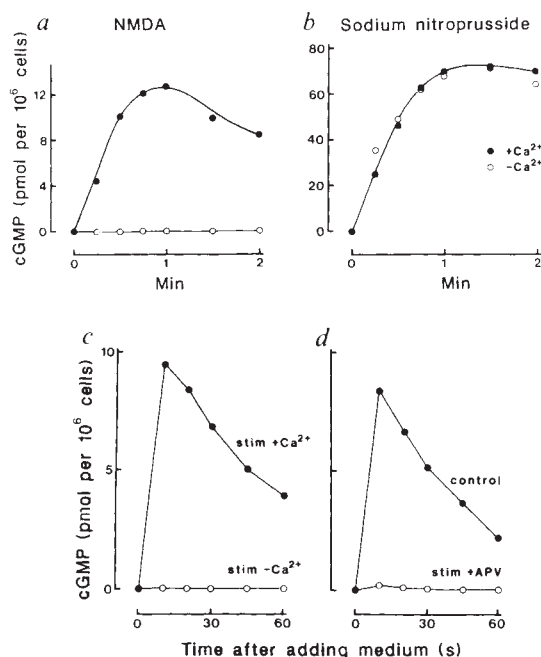


Fig. 4 *a* and *b*, Differential dependence on Ca²⁺ of cGMP responses to NMDA (*a*) and sodium nitroprusside (*b*) in cerebellar cell suspensions. The cells were preincubated in the presence of 1.5 mM Ca²⁺. Aliquots of the suspension were then exposed to 100 μ M NMDA or 300 μ M sodium nitroprusside in the absence (●) or presence (○) of EGTA in a concentration (4 mM) calculated to reduce the free Ca²⁺ concentration to 40 nM without markedly affecting the free Mg²⁺ concentration (calculated to be 1 mM at 1.2 mM total Mg²⁺). The EGTA was added 1 min before the NMDA or sodium nitroprusside. *c* and *d*, NMDA-induced release of EDRF from cerebellar cells is Ca²⁺-dependent and inhibited by APV. The protocol was based on the one described in Fig. 2. In (*c*), cells incubated with 1.5 mM Ca²⁺ and superoxide dismutase (30 U ml⁻¹) were stimulated with NMDA (100 μ M) for 30 s. EGTA (4 mM) was added either 25 s after (●) or 1 min before (○) the NMDA. The suspending medium was removed and 0.5 ml of it added (at *t*=0) to 0.5 ml detector cells containing D-APV (100 μ M), superoxide dismutase (30 U ml⁻¹) and EGTA (4 mM). cGMP levels in the detector cells were determined at the intervals indicated. *d*, The detector cells contained no EGTA (that is, 1.5 mM Ca²⁺) and the exposure to NMDA was carried out in the absence (●) or presence (○) of D-APV (50 μ M).

ing that it is unstable. A more direct test of its stability (Fig. 2c) showed that the activity of the medium in the absence of superoxide dismutase declined with a half life of about 18 seconds. This matches the half life of EDRF under similar conditions ($pO_2 \sim 150$ mm Hg; 37 °C)¹³. In the presence of superoxide dismutase, the half life increased about twofold, which is consistent with the effect of superoxide dismutase on the stability of EDRF⁷.

If the factor released following NMDA receptor activation is EDRF, it should relax vascular smooth muscle. This was tested using opened rings of rat aorta maintained *in vitro* and contracted with 100 nM phenylephrine. NMDA (100 μ M) had no effect on muscle tension when applied in normal conditions (Fig. 3a). When cerebellar cells were included in the bathing medium, however, NMDA elicited a marked relaxation in each muscle preparation tested ($n=5$). The NMDA-induced relaxations were inhibited by haemoglobin and by APV. The contracting agent used did not modify cGMP responses to NMDA in cerebellar cell suspensions (Fig. 3b) and so is unlikely to have affected the release of the factor.

On the basis of these results, we propose that NMDA receptor activation leads to the release of EDRF and that EDRF is responsible for the associated increases in cGMP in the cerebel-

lum. In addition, the inhibitory effect of haemoglobin, a large protein unable to penetrate intact cells, suggests that accumulation of EDRF in the extracellular fluid is a necessary step in the cGMP response in cerebellar cell suspensions. This indicates that the principal targets of EDRF are unlikely to be the cells that release it, a conclusion fully compatible with previous results⁵. In the young rat cerebellum, lesion experiments coupled with studies of the location of sodium nitroprusside-activated guanylate cyclase point to the neurons, and especially the granule cells, as the main source of EDRF and the glia as the main targets^{5,14}. In glia, one effect of EDRF (through cGMP) may be on membrane K⁺ or Ca²⁺ conductances¹⁵. Preparations enriched in presynaptic nerve terminals from the cerebellum¹⁴ and elsewhere¹⁶ are also rich in sodium nitroprusside-stimulated guanylate cyclase activity and so these elements must be regarded as another likely target. Any function of cGMP generated presynaptically in response to EDRF remains to be studied, although retrograde transsynaptic factors have been invoked to explain several phenomena, one example being long-term potentiation in the hippocampus. This is a form of synaptic plasticity whose induction is believed to involve Ca²⁺-influx through NMDA-gated ion channels¹⁷.

In this regard, cGMP responses to NMDA in the cerebellar cell suspensions were dependent on extracellular Ca²⁺ (Fig. 4a) whereas responses to sodium nitroprusside were not (Fig. 4b), indicating that Ca²⁺ may be involved in the initial synthesis/release of EDRF. To test this, cerebellar cells were exposed to NMDA in the presence or absence of 4 mM EGTA ([Ca²⁺]₀ = 40 nM or 1.5 mM). After 30 seconds, the suspending medium was removed and added to detector cells as before. With normal Ca²⁺ in the initial exposing solution, the usual transient cGMP elevation in the detector cells took place irrespective of whether the detector cells were incubated with or without EGTA (Fig. 4c, d). Once released, therefore, the action of the factor, like that of sodium nitroprusside, is Ca²⁺-independent. With EGTA present during the initial exposure to NMDA, however, the medium was subsequently unable to produce a cGMP response in detector cells (Fig. 4c). Thus, as in the vascular system¹⁸, the release of EDRF is the Ca²⁺-dependent step. The ion channels linked to NMDA receptors have a high Ca²⁺-permeability¹⁹ and cGMP responses to NMDA are sensitive to Mg²⁺ (ref. 20), an ion which blocks these channels²¹, so influx of Ca²⁺ through NMDA-linked channels represents a likely trigger for EDRF production.

As to the immediate source, L-arginine has recently been proposed to be the precursor of EDRF in endothelial cells, NO being derived from the guanidino group of this amino acid²². L-arginine was in fact identified several years ago as the compound present in rat brain which, in the presence of other unidentified factors, produced a sodium nitroprusside/NO-like activation of guanylate cyclase²³. It seems probable that neurons in the CNS and vascular endothelial cells synthesize EDRF by the same mechanism.

The work was supported by the MRC.

Received 9 June; accepted 10 October 1988.

1. Furchgott, R. F. & Zawadzki, J. V. *Nature* **288**, 373-376 (1980).
2. Vanhoutte, P. M., Rubanyi, G. M., Miller, V. M. & Houston, D. S. A. *Rev. Physiol.* **48**, 307-320 (1986).
3. Waldman, S. A. & Murad, F. *Pharmac. Rev.* **39**, 163-196 (1987).
4. Drummond, G. I. *Adv. Cyclic Nucleotide Res.* **15**, 373-494 (1983).
5. Garthwaite, J. & Garthwaite, G. *J. Neurochem.* **48**, 29-39 (1987).
6. Garthwaite, J. *Br. J. Pharmac.* **85**, 297-307 (1985).
7. Rubanyi, G. M. & Vanhoutte, P. M. *Am. J. Physiol.* **250**, H815-H821 (1986).
8. Gryglewski, R. J., Palmer, R. M. J. & Moncada, S. *Nature* **320**, 454-456 (1986).
9. Palmer, R. M. J., Ferrige, A. G. & Moncada, S. *Nature* **327**, 524-526 (1987).
10. Ignarro, L. J., Byrns, R. E., Buga, G. M. & Wood, K. S. *Circulation Res.* **61**, 866-879 (1987).
11. Martin, W., Villani, G. M., Jothianandan, D. & Furchgott, R. F. *J. Pharmac. exp. Ther.* **232**, 708-716 (1985).
12. Gruetter, C. A., Kadowitz, P. J. & Ignarro, L. J. *Can. J. Physiol. Pharmac.* **59**, 150-156 (1981).
13. Forstermann, U., Trogisch, G. & Busse, R. *Eur. J. Pharmac.* **106**, 639-643 (1985).
14. Bunn, S. J., Garthwaite, J. & Wilkin, G. P. *Neurochem. Int.* **8**, 179-185 (1986).
15. Krnjevic, K. & Van Meter, W. G. *Can. J. Physiol. Pharmac.* **54**, 416-421 (1976).
16. Nakane, M. & Deguchi, T. *Biochim. biophys. Acta* **525**, 275-285 (1978).
17. Collingridge, G. L. & Bliss, T. V. P. *Trends Neurosci.* **10**, 288-293 (1987).

18. Peach, M. J., Singer, H. A., Izzo, N. J. Jr & Loeb, A. L. *Am. J. Cardiol.* **59**, 35A-43A (1987).
19. MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J. & Barker, J. L. *Nature* **321**, 519-522 (1986).
20. Garthwaite, J. *Neuroscience* **7**, 2491-2497 (1982).
21. Nowak, L., Bregestovski, P., Ascher, P., Herbert, A. & Prochiantz, A. *Nature* **307**, 462-465 (1984).
22. Palmer, R. M. J., Ashton, D. S. & Moncada, S. *Nature* **333**, 664-666 (1988).
23. Deguchi, T. & Yoshioka, M. *J. biol. Chem.* **257**, 10147-10151 (1982).

Deletion of self-reactive T cells before entry into the thymus medulla

Hans Hengartner*, Bernhard Odermatt*,
Reto Schneider*, Magali Schreyer†, Gaby Wälle*,
H. Robson MacDonald† & Rolf M. Zinkernagel*

* Institute of Pathology, University Hospital, CH-8091 Zürich, Switzerland

† Ludwig Institute for Cancer Research, Lausanne Branch, CH-1066 Epalinges, Switzerland

The thymus is important in the differentiation of bone marrow-derived precursor cells into functional T cells; humoral factors¹⁻⁴, as well as physical interactions with nurse cells⁵, dendritic cells and epithelial cells⁶⁻⁸, are thought to be instrumental in this process. Thymic lymphocytes mature during their migration from the cortical to the medullary region of the thymus⁹⁻¹¹, when they undergo phenotypic changes that include the acquisition of T-cell antigen receptors^{12,13}, hormone receptors^{1,14,15} and differentiation antigens¹⁶. Cortical T cells are thus mostly CD4⁺CD8⁺, whereas medullary T cells are either CD4⁺CD8⁻ or CD4⁻CD8⁺ (refs 6, 16 and 17). During this period T cells are subjected to two types of repertoire selection: all T cells recognizing self-MHC with low affinity¹⁸ may be preferentially amplified (positive selection), and in a second step T cells with high-affinity receptors for self-MHC determinants plus self antigens are eliminated (negative selection). We have described two monoclonal antibodies¹⁹ specific for the Vβ6 gene segment of the α/β heterodimeric T-cell antigen receptor and have shown that most CD4⁺Vβ6⁺ T cells recognize the Mls^a antigenic determinant but not Mls^b (ref. 20; similar results have been reported for Vβ8.1 and Mls^a (ref. 21). In both situations, tolerance to Mls^a correlated in an MHC-dependent fashion with absence of Vβ6 or Vβ8.1 T-cell antigen receptor expressing T cells in the periphery. We show here by immunostaining of thymus cryosections and cytofluorometric analysis that Vβ6-expressing cortical T cells are present at high density in both Mls^a and Mls^b mice, but do not enter the medullary region of Mls^a animals.

Thymus cryosections, as well as thymus and lymph node cell suspensions from BALB/c (H-2^d, Mls^b), DBA/2 (H-2^d, Mls^a), BALB/c.D2.Mls^a (H-2^d, Mls^a) and SJL (H-2^s, Mls^c) mice were stained with the rat monoclonal antibody (mAb) 44-22-1 (ref. 19) (anti-Vβ6) and KJ16-133 (ref. 22) (anti-Vβ8) as a control antibody. Cells reacting with the antibody KJ16-133 (anti-Vβ8.1 and anti-Vβ8.2) (ref. 23) were localized as weakly staining in the cortical region and as brightly staining T cells in the medullary area, confirming earlier immune histological and FACS analyses^{10,24}. Thymuses of SJL mice which lacked the Vβ8 genes²⁵ showed no positive cells.

The immunostaining of thymus cryosections of all the indicated mouse strains with the Vβ6-specific mAb 44-22-1 showed the usual weak staining of cortical thymocytes. The medullary thymocytes of Mls^a-animals, however, remained unstained (Fig. 1c and e), whereas those of Mls^b-animals stained as intensely as with KJ16-133 (Fig. 1a and g). These observations were quantitated morphometrically (Table 1) and confirmed by flow-microfluorometric analysis of thymus and lymph node T cells using the same antibodies (Table 2, and representative FACS-histograms in Fig. 2).

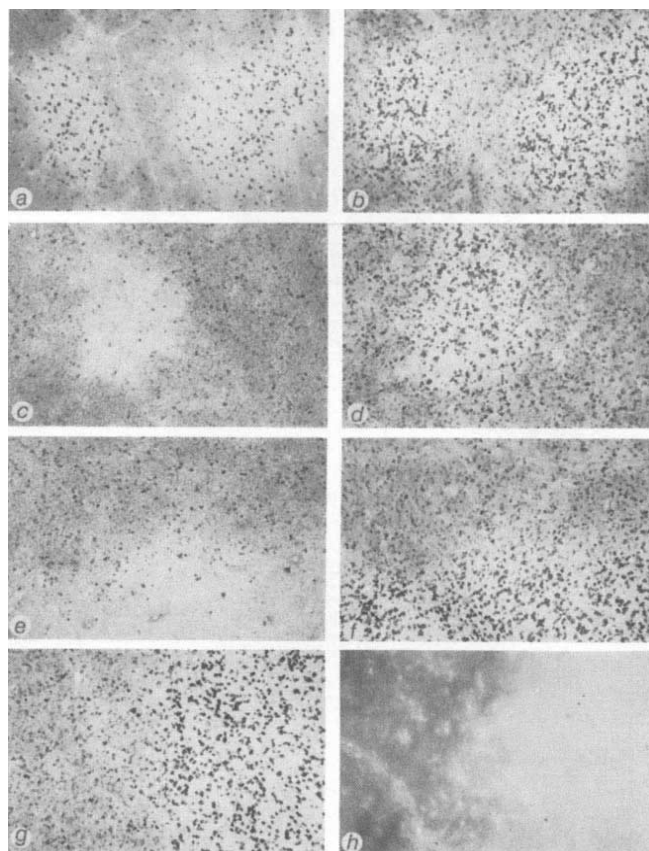


Fig. 1 Expression of Vβ6 and Vβ8 TCR on cortical (darker area) and medullary (lighter area) thymocytes. Sequential cryosections of mouse thymuses were stained with rat mAb 44-22-1 (anti-Vβ6; a, c, e and g) or KJ16-133 (anti-Vβ8; b, d, f and h). a and b, BALB/c (H-2^d, Mls^d); c and d, DBA/2 (H-2^d, Mls^a); e and f, BALB/c.D2.Mls^a (H-2^d, Mls^a); g and h, SJL (H-2^s, Mls^c).

Methods. Freshly removed thymuses were immersed in Hank's balanced salt solution and snap-frozen in liquid nitrogen. Tissue sections of 5 μm thickness were cut in a cryostat and immunostained (H. Stein *et al.*, personal communication). Sequential sections were fixed in acetone for 10 min and incubated with undiluted hybridoma culture supernatant containing either rat mAb 44-22-1 (anti-Vβ6) or rat mAb KJ16-133 (anti-Vβ8) for 30 min. Affinity-purified peroxidase-labelled goat antibodies to rat immunoglobulins (Jackson Laboratories, Pennsylvania) were then diluted by 1 in 40, applied for 30 min and followed by peroxidase-labelled pig antibodies to goat immunoglobulins (Tago) diluted 1 in 20 for another 30 min. Dilutions of antisera were made in phosphate-buffered saline containing 5% normal mouse serum. Peroxidase was then detected by the Graham-Karnowsky reagent (R. C. Graham and M. J. Karnouallis, personal communication). Sections were counterstained with Meyer's hemalum for 2 min. The signal was amplified if necessary by osmification with 1% osmium tetroxide in Tris-buffered saline for 15 min. Photographs were taken with a Leitz Orthoplan microscope using Kodak TP-135 technical film. Magnification, ×135.

The direct demonstration that Vβ6-expressing (potentially self-reactive) T lymphocytes are absent in the thymus medulla, but present at normal levels in the cortex of Mls^a mice is consistent with the possibility that elimination occurs at the cortico-medullary junction. An alternative but not mutually exclusive interpretation of our data is that elimination of self-reactive T cells occurs simultaneously at multiple sites within the thymus cortex.

Tolerance induction in the cortico-medullary junction may be mediated by dendritic cells, which are found at high density in this region²⁶. It has been shown that thymic dendritic cells

Fig. 2 Flow microfluorometric estimations of V β 6 and V β 8 TCR densities on thymocytes. Numbers of cells in arbitrary units. Thymocytes from BALB.D2.Mls^a (H-2^d, Mls^a) and BALB/c (H-2^d, Mls^b) animals were stained with anti-V β 6 mAb 44-22-1 (Fig. 2a and b) and thymocytes from SJL (H-2^s, Mls^c) and BALB/c (H-2^d, Mls^b) animals with anti-V β 8 mAb KJ16-133 (Fig. 2c and d) followed by fluoresceinated goat anti-rat immunoglobulin. Thin lines represent the fluorescent conjugate alone. Materials and methods are described in the legend to Table 1.

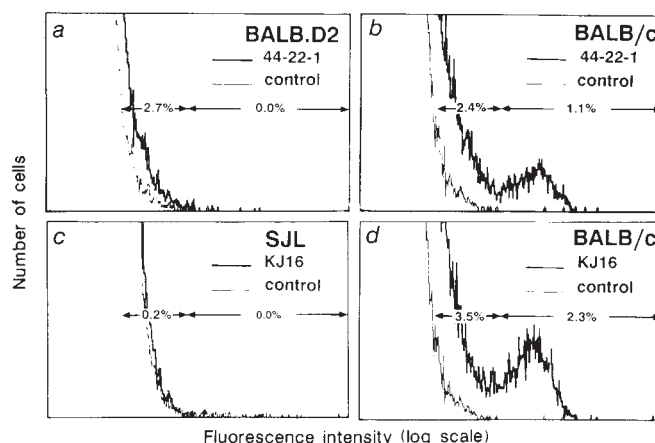


Table 1 Morphometric analysis of anti-V β 6 and anti-V β 8 stained thymus sections

Strain	Cortex						Medulla			
			KJ16-133		44-22-1		KJ16-133		44-22-1	
	H-2	Mls	weak	strong	weak	strong	weak	strong	weak	strong
BALB/c	d	b	148 ± 9	2 ± 1	63 ± 6	<1	8 ± 6	145 ± 15	5 ± 2	65 ± 7
DBA/2	d	a	127 ± 18	5 ± 2	66 ± 8	2 ± 1	4 ± 3	143 ± 14	1 ± 1	1 ± 1
BALB/c.D2.Mls ^a	d	a	139 ± 19	3 ± 2	59 ± 8	2 ± 1	5 ± 2	162 ± 14	1 ± 1	1 ± 1
SJL	s	c	<1	<1	82 ± 11	2 ± 1	<1	<1	2 ± 1	135 ± 14

Thymus cryosections from untreated H-2^d congenic and control mouse strains were indirectly stained by an immunoperoxidase method with mAb KJ16-133 (V β 8), or mAb 44-22-1 (V β 6). Strongly and weakly stained cells in the thymic cortex and medulla were counted separately. The mean number of stained cells per test area with an average total thymocyte number of 600 in each category was determined. Each number represents the mean of 10 test areas. Cryosections were stained as described in Fig. 1. Cells were divided by subjective criteria into weakly and strongly stained populations. Cells were counted using a microscope equipped with the MOP-AMO3 device (Kontron, Zürich) in test areas belonging entirely either to the cortical or medullary regions.

Table 2 Expression of V β 8 and V β 6 by T-cells from lymph nodes and thymus

Strain	Lymph nodes				Thymus			
			KJ16 (V β 8)		KJ16 (V β 8)		44-22-1 (V β 6)	
	H-2	Mls	weak	strong	weak	strong	weak	strong
BALB/c	d	b	14.0	7.5	3.5	2.3	2.4	1.1
BALB.D2.Mls ^a	d	a	13.6	0.4	4.0	1.6	2.7	0.0
(BALB/c × BALB.D2.Mls ^a)F1	d × d	b × a	11.4	0.4	3.8	0.9	3.7	0.1
DBA/2	d	a	8.7	0.7	4.3	2.0	2.2	0.1
SJL	s	c	0.1	8.1	0.2	0.0	3.3	1.8

Lymph node cells and untreated thymocytes from H-2^d congenic and control mouse strains (SJL) were stained by indirect immunofluorescence with mAb KJ16 (V β 8), or mAb 44-22-1 (V β 6), followed by FITC-conjugated goat anti-rat-IgG and analysed by flow microfluorometry. Data are presented as % positive cells after subtraction of staining with the fluorescent second antibody alone. Aliquots of 1–5 × 10⁶ cells in 100 μ l were stained at 4 °C as described³². Antibodies 44-22-1 (ref. 19) and KJ16 (ref. 22) were used as undiluted hybridoma culture supernatants. Goat-anti-rat immunoglobulin (absorbed on mouse immunoglobulin) was from Tago, Burlingame, California. Stained samples were passed on a EPICS profile flow cytometer (Coulter Electronics) gated to exclude non-viable cells. Histograms represent 5 × 10⁴ viable thymocytes, whereas only 10⁴ lymph node cells were usually counted. The proportion of weakly staining thymocytes was underestimated because of overlap with the negative control.

can efficiently present intravenously administered protein antigens⁸ and there is good evidence that dendritic cells are the main cell type able to transfer antigen-specific tolerance in radiation bone marrow chimaeras²⁷. Also, isolation of complexes of thymocyte and stromal cells has shown that an unexpectedly high proportion of CD4⁺CD8⁺ thymocytes were tightly associated with medullary dendritic cells²⁸. The occurrence of clonal deletion of V β 6⁺ T cells at the CD4⁺CD8⁺ precursor-T cell stage^{29,30} and the data presented here suggest that CD4⁺CD8⁺ thymocytes encountering putatively tolerogenic dendritic cells at the cortico-medullary interface might be deleted. Whether this deletion is an active process, or simply induction of a developmentally regulated suicide programme³¹ cannot be inferred from morphological observations.

We thank M. Condrau for FACS-analyses. This work was supported in part by the Swiss National Science Foundation (H.H.) and the Radium Stiftung Zürich.

Received 9 September; accepted 25 October 1988.

- Howe, R. C., Lowenthal, J. W. & MacDonald, H. R. *J. Immun.* **137**, 3195–3200 (1986).
- Palacios, R., Sideras, P. & von Boehmer, H. *EMBO J.* **6**, 91–95 (1987).
- Goldstein, A. L. & Low, T. L. K. *Prog. Immun.* **5**, 1417–1427 (1983).
- Bach, J. F. *Prog. Immun.* **5**, 1563–1579 (1983).
- Wekerle, H., Ketelsen, U. P. & Ernst, M. *J. exp. Med.* **151**, 925–944 (1980).
- Longo, D. L. & Schwartz, R. H. *Nature* **287**, 44–45 (1980).
- Kyewski, B. A., Fathman, C. G. & Rouse, R. V. *J. exp. Med.* **163**, 231–246 (1986).
- Kyewski, B. A., Fathman, C. G. & Kaplan, H. S. *Nature* **308**, 196–199 (1984).
- Kingston, R., Jenkinson, E. J. & Owen, J. J. T. *Nature* **317**, 811–813 (1985).
- Farr, A. G., Anderson, S. K., Marrack, P. & Kappler, J. W. *Cell* **43**, 543–550 (1985).
- Weissman, I. L. *J. exp. Med.* **126**, 291–304 (1967).
- Budd, R. C. *et al. J. exp. Med.* **166**, 577–582 (1987).

13. von Boehmer, H., Karjalainen, K., Pelkonen, J., Borgulya, P. & Rammensee, H. G. *Immunol. Rev.* **101**, 21–37 (1988).
14. Lugo, J. P. *et al. J. exp. Med.* **161**, 1048–1062 (1985).
15. Lowenthal, J. W., Howe, R. C., Ceredig, R. & MacDonald, H. R. *J. Immun.* **137**, 2579–2584 (1986).
16. Scollay, R., Bartlett, P. & Shortman, K. *Immunol. Rev.* **82**, 79–103 (1984).
17. Crispe, I. N., Moore, M. W., Husmann, L. A., Smith, L. & Bevan, M. J. *Nature* **329**, 336–339 (1987).
18. Zinkernagel, R. M. *et al. J. exp. Med.* **147**, 882–896 (1978).
19. Payne, J. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
20. MacDonald, H. R. *et al. Nature* **332**, 40–45 (1988).
21. Kappler, J. W., Staerz, U. D., White, J. & Marrack, P. *Nature* **332**, 35–40 (1988).
22. Haskins, K. *et al. J. exp. Med.* **157**, 1149–1169 (1984).
23. Behlke, M. A. *et al. J. exp. Med.* **165**, 257–262 (1987).
24. Cristanti, A., Colantoni, A., Snodgrass, R. & von Boehmer, H. *EMBO J.* **5**, 2837–2843 (1986).
25. Sim, G. K. & Augustin, A. A. *Cell* **42**, 89–92 (1985).
26. Barclay, A. N. & Mayrhofer, G. *J. exp. Med.* **153**, 1666–1671 (1981).
27. Sprent, J., Lo, D., Gao, K. E. & Ron, Y. *Immunol. Rev.* **101**, 173–190 (1988).
28. Kyewski, B. A., Momburg, F. & Schirmacher, V. *Eur. J. Immun.* **17**, 961–967 (1987).
29. MacDonald, H. R., Hengartner, H. & Pedrazzini, T. *Nature* **335**, 174–176 (1988).
30. Fowlkes, B. J., Schwartz, R. H. & Pardoll, D. N. *Nature* **334**, 620–623 (1988).
31. Duvall, E. & Wyllie, A. H. *Immun. Today* **7**, 115–119 (1986).
32. MacDonald, H. R., Blanc, C., Lees, R. L. & Sordat, B. *J. Immun.* **136**, 4337–4339 (1986).

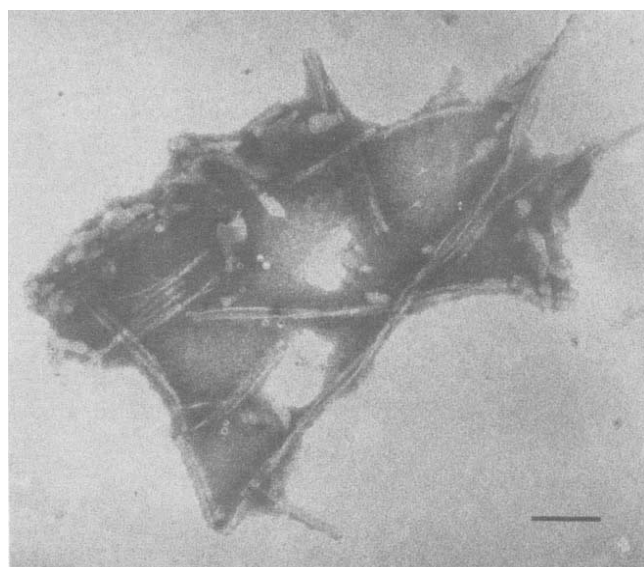


Fig. 1 Electron micrograph of negatively-stained fibrils from the spinal cord (C1/C2 segments) of a BSE-affected cow.

Methods. Four brains were parasagittally bisected from cows in which clinical signs had persisted for 4–6 months. The larger left portion was fixed in 10% formal saline for histopathological analysis. The remaining tissue was dissected into the following anatomic areas and frozen at -70°C : basal ganglia, frontal, parietal and occipital lobes of cerebrum, cerebellum, medulla (caudal to the obex) and spinal cord segments C1 and C2. To prepare BSE fibrils, tissues were thawed, homogenized in 10% (w/v) *N*-lauroylsarcosine and processed as previously described⁵. BSE fibrils were stained on formvar/carbon grids for electron microscopy with 2% (w/v) potassium phosphotungstate, pH 6.6 and grids examined in a Philips EM 410LS electron microscope at 80 kV under $\times 31,000$ magnification. Each sample grid was examined for a minimum of 20 min and assessed to have no fibrils, some fibrils or many fibrils. Typical BSE fibrils are illustrated; scale bar, 100 nm.

Fibrils from brains of cows with new cattle disease contain scrapie-associated protein

James Hope*, Laura J. D. Reekie*, Nora Hunter*,
Gerd Multhaupt†, Konrad Beyreuther†, Heather White‡,
Anthony C. Scott‡, Michael J. Stack‡, Michael Dawson‡
& Gerald A. H. Wells‡

* Institute for Animal Health, AFRC & MRC Neuropathogenesis Unit, Ovgston Building, West Mains Road, Edinburgh EH9 3JF, UK

† Zentrum für Molekulare Biologie, Universität Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, FRG

‡ Ministry of Agriculture, Fisheries and Food, Central Veterinary Laboratory, New Haw, Weybridge, Surrey KT15 3NB, UK

During the past two years, more than 1,000 cases of a neurological disorder of cattle, bovine spongiform encephalopathy (BSE)¹, have been confirmed from farms throughout Great Britain. The neurological signs and brain pathology of BSE resemble those produced in other species by the pathogens of scrapie and related disorders. The discovery of fibrils similar to scrapie-associated fibrils in detergent extracts of BSE-affected brain supported the clinical and pathological diagnosis of the disease¹, but has been controversial. Scrapie-associated fibrils are found in brain extracts of all species affected by scrapie and diseases caused by related pathogens^{2,3}. They are pathological aggregates of a neuronal membrane protein termed PrP^{4–6} and a protease-resistant form of PrP^{7–9} is a molecular marker of scrapie-associated fibrils. In this report, we show the major protein of BSE fibrils is the bovine homologue of PrP as judged by its size, protease resistance, immunoreactivity, lectin binding and partial N-terminal protein sequence. This confirms that BSE is a scrapie-like disease.

To determine if BSE fibrils are the bovine equivalent of scrapie-associated fibrils (SAF) we needed to obtain sufficient fibril protein for chemical characterization and protein sequencing. The amounts of SAF in brains of sheep affected by natural scrapie are highly variable^{10,11} and, in small rodent models of the disease, depend on the area of brain sampled amongst other factors¹². To optimize our purification work, we screened areas of BSE-affected cow brain for the distribution of BSE fibrils. All suspected BSE cases were confirmed by histological examination before purification of fibrils from contralateral or adjacent brain areas (in natural scrapie or BSE, the pathognomic vacuolar lesions have a bilaterally-symmetrical distribution¹). Comparisons were made of the estimated yields of BSE fibrils purified from each brain area with the severity of its vacuolation. Severe vacuolation of neurons and grey matter neuropil was

found in the brain stem in each case. The caudal medulla and cervical spinal cord areas had moderately severe vacuolation. Vacuolation was mild in the basal ganglia, the occipital lobe of the cerebrum (localized to the hippocampus) and the cerebellum (localized mainly to the central parts of the vermis). The distribution of BSE fibrils was consistent in each brain but varied between brain areas. Large numbers were found in the basal ganglia. A few fibrils were found in the other brain areas although they were not seen in occipital lobe preparations in two cases. This variation in yield of fibrils may reflect differences in the extraction and purification efficiencies of our methods when applied to areas of the brain which differ in their relative amounts of grey and white matter. Alternatively, the yield of BSE fibrils may indicate the degree of agent replication in that area of the brain, the level of tissue damage, or both. In this survey, no correlation could be drawn between the occurrence or severity of vacuolation and the number of BSE fibrils purified by detergent extraction of contralateral anatomic brain areas. Previous work using whole brain or selected areas of cortical grey matter from experimental rodent¹² and sheep scrapie¹⁰, or natural scrapie^{10,11}, has also shown SAF are not markers for vacuolation of neurons but are probably indicative of a more fundamental pathological process.

Typical BSE fibrils are shown in Fig. 1, and have the size and shape of sheep SAF purified in our laboratories (data not shown). The morphology of these aggregates depends on the methods used for their purification; in fact some drastic treatments may yield aggregates of protein that are unrecognizable as SAF¹³. Differences in procedure may underlie the initial

Fig. 2 The proteins of the disease-specific fibrils of BSE and natural scrapie. Lanes 1-7, BSE-fibril proteins (0.15 g equivalents of cattle brain); lanes 8-11, SAF proteins (0.1 g equivalents of sheep brain): 1, spinal cord segments, C1/C2 (-PK); 2, spinal cord segments, C1/C2 (+PK); 3, occipital cortex (-PK); 4, parietal cortex (-PK); 5, frontal cortex (-PK); 6, basal ganglia (-PK); 7, basal ganglia (+PK); 8, sheep SAF (-PK); 9, sheep SAF (+PK); 10, control (-PK); 11, control (+PK). The major protein of BSE fibrils has a M_r of 33-35K (PrP 33-35), similar to that found in hamster and mouse SAF^{5,19}. A minor variant of M_r 26-29K (PrP 26-29) was also seen (lanes 1, 3-6). On proteinase K hydrolysis, PrP 33-35 is converted to a protease-resistant fragment of M_r 27-30K (PrP 27-30), lanes 2, 7, whereas PrP 26-29 is reduced in M_r to 23-25K and can barely be seen because of its low abundance (lanes 2, 7). The similar staining intensity of PrP 26-29 (lane 1) and PrP 27-30 (lane 2) is coincidental; these relative molecular mass variants can be more readily distinguished by two-dimensional gel analysis. These data were consistent with results from three other BSE-affected cow brains.

Methods. BSE fibrils or SAF were prepared with (+PK) or without (-PK) the use of proteinase K from different brain areas of a BSE-affected Holstein/Friesian cow or the cerebral cortex of a Cheviot sheep affected by natural scrapie. Equivalent fractions were prepared from an unaffected sheep. Fractions were analysed by SDS-PAGE and proteins visualized by silver staining.

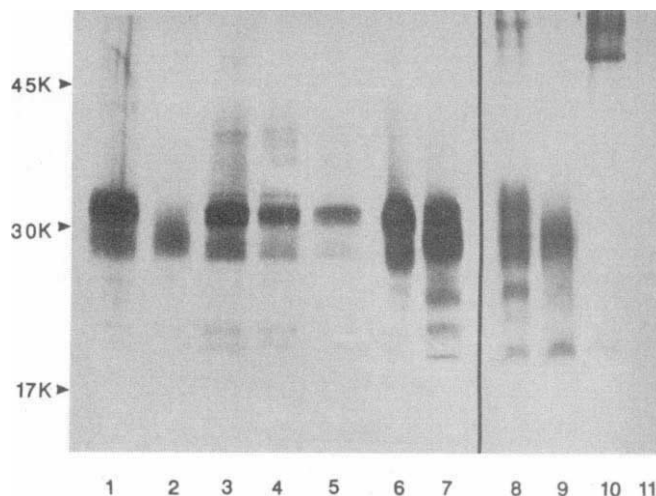
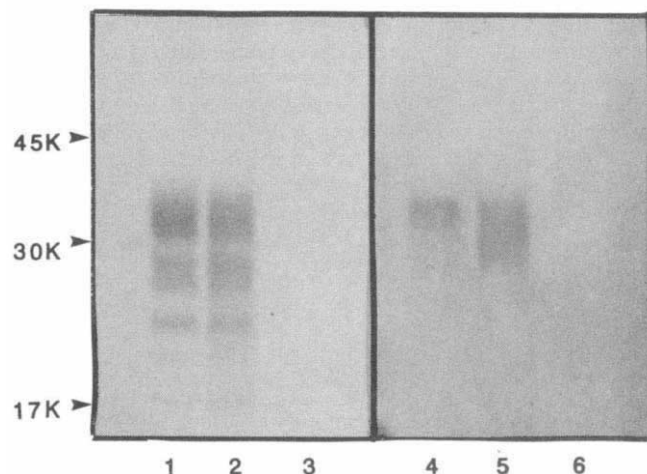


Fig. 3 Detection of PrP-related glycoprotein in BSE fibrils by lectin and immunoblotting. Lanes 1, 2 and 4, BSE-fibril fractions purified without proteinase K; lane 5, BSE-fibril fraction partially hydrolysed with proteinase K; lanes 3, 6, equivalent fractions from unaffected cow brain.

Methods. BSE fibrils were prepared from basal ganglia of a BSE-affected cow (see Fig. 1 legend). Equivalent fractions were prepared from the basal ganglia of an unaffected cow. Fractions were analysed by SDS-PAGE, electroblotted onto nitrocellulose and specific proteins detected by either (1) a rabbit anti-mouse SAF serum (final dilution, 1 in 1,000; Farquhar and Somerville, manuscript in preparation) and a gold-conjugated goat anti-rabbit immunoglobulin/silver enhancement system (Janssen) (lanes 1-3; 0.1 g equivalents of bovine brain) or (2) biotinylated wheat germ agglutinin (Vector Labs) (lanes 4-6; 0.24 g equivalents of bovine brain). Biotinylated lectins were visualized using alkaline-phosphate-conjugated streptavidin¹⁸. No proteins were visualized if the rabbit anti-mouse SAF serum was replaced by non-immune rabbit serum. Other biotinylated lectins (ricin, lentil lectin and concanavalin A) also bound to one or more of the M_r variants of BSE-fibril protein (data not presented). Similar immunochemical and lectin binding data were obtained from other brain areas of the same cow and the brains of three additional cases of BSE.



confusion over the identity of SAF and the rod structures originally described as co-purifying with PrP protein; as SAF protein and PrP protein seems to be chemically equivalent molecules^{5,14,15}, the implication is that PrP rods and SAF are different morphological forms with a common protein subunit.

The protein composition of BSE fibrils was determined by SDS-PAGE and silver-stain analysis. Fibril fractions prepared without proteases contained a major protein of relative molecular mass (M_r) 33,000-35,000 (33K-35K) and a minor (26-29K) variant, and proteinase K treatment produced a 27-30K core protein (Fig. 2). The yield of BSF fibril proteins was estimated by inspection of silver-stained protein bands on SDS-PAGE gels using at least two different concentrations. The ranking of brain areas for amounts of protein correlated with the yield of BSE-fibrils seen in comparable fractions by electron microscopy: basal ganglia > medulla > spinal cord > cerebellum and cortex.

How do these properties of the BSE fibril protein compare with PrP in SAF? PrP is a neuronal protein of apparent M_r 33-35K (PrP33-35)^{5,15,16}. It has two potential N-glycan attachment sites and an N-terminal, glycine-rich repetitive sequence¹⁵. The scrapie-induced aggregation of PrP into fibrils partially

protects the protein from proteolysis, and PrP was originally purified as a 27-30K polypeptide (PrP27-30) using detergents and proteinase K^{7-9,14,15}. Proteinase K cleaves with the glycine-rich repetitive region of the protein in SAF⁵. Under these conditions, PrP in uninfected brain can be solubilized from membranes and completely hydrolysed¹⁶. Our chemical analyses indicate that the BSE fibril protein has an apparent M_r of 33-35K and, in BSE fibrils, is partially resistant to proteolysis with proteinase K. The protease-resistant core of BSE fibril protein has an apparent M_r of 27-30K. The BSE fibril proteins were similar in size and protease sensitivity to PrP in SAF purified from the brain of a Cheviot sheep with natural scrapie (Fig. 2a, lanes 8, 9). The putative bovine PrP protein also cross-reacted with anti-mouse SAF protein antisera (Fig. 3, lanes 1, 2) and bound specifically to lectins, indicating that it is glycosylated (Fig. 3, lanes 4, 5). The hamster SAF protein is glycosylated¹⁷, and the heterogeneity of relative molecular mass of PrP in SAF (and of BSE fibril protein) may be due to differences in the carbohydrate structures of this glycoprotein¹⁷⁻¹⁹. This protein and its molecular size variants were not found in corresponding fractions of uninfected brain (Fig. 3, lanes 3, 6).

Further evidence that BSE fibrils are composed of the bovine

Fig. 4 N-terminal sequences of SAF- and BSE-fibril proteins. The BSE fibril protein has the same N-terminal amino-acid sequence as ovine PrP and has an extra glycine residue compared with the N-terminal structure of human and rodent PrP^{5,19-21}. BSE-fibril protein was purified and sequenced by published methods.⁵.

Lys-Lys-Arg-Pro-Lys-Pro-Gly-Gly-Trp-Asn-Thr												Mouse, Hamster Human
1				5							10	
Lys-Lys-Arg-Pro-Lys-Pro-Gly-Gly-Gly-Trp-Asn-Thr												Cattle, Sheep
1				5							10	

PrP homologue came from amino-acid sequencing (Fig. 4). Purification of fibrils from BSE-affected brain cortex (20 g) gave low, but sufficient, amounts of protein for sequence determination (10 pmol). Its N-terminal 12 amino acids were identical to those found in sheep SAF protein (Goldmann *et al.*, manuscript in preparation) and differed from mouse¹⁹, hamster⁵, and human^{20,21} PrP by one amino acid. The bovine and sheep homologues have an extra glycine residue inserted between residues 9 and 10 of the rodent and human sequences (Fig. 4). Whether this mutation of PrP is needed for natural transmission of these sub-viral pathogens remains obscure.

PrP is the protein product of a highly conserved gene found in species as diverse as man²¹, cattle²² and fruit fly²³. The protein has a key role in the pathogenesis of scrapie and related natural and experimental disorders: it is the major component of SAF⁴⁻⁶ and amyloid-containing neuritic plaques^{24,25}; it co-purifies with infectivity¹⁴ and may be a host-component of the infectious unit of scrapie²⁶. Additionally, the close linkage of the PrP gene to the murine gene controlling scrapie incubation times (*Sinc*)²⁷⁻²⁹ indicates that it may be part of a host defence mechanism against these unusual pathogens. The protein may mediate its effects on the development of these neurodegenerative diseases by acting as a cellular receptor for the pathogen³⁰. Our results indicate that the bovine PrP homologue has a similar role in the pathogenesis of BSE.

BSE probably originated in cattle which were inadvertently fed sheep meat and bone meal contaminated with scrapie³¹. Our protein data and the recent production of a scrapie-like disease in mice injected with BSE-affected cow brain³² confirm the homology of this novel cattle disorder and scrapie.

Financial support for protein sequencing was obtained through the Deutsche Forschungsgemeinschaft, the Bundesministerium für Forschung und Technologie, and the Fonds der chemischen Industrie of the Federal Republic of Germany. The progress of this work benefited by the award of a NATO collaborative research grant to J.H. and K.B.

Received 2 June; accepted 31 October 1988.

- Wells, G. A. H. *et al. Vet. Rec.* **121**, 419-420 (1987).
- Merz, P. A., Somerville, R. A., Wisniewski, H. M. & Iqbal, K. *Acta neuropath.* **65**, 63-74 (1981).
- Merz, P. A., Somerville, R. A., Wisniewski, H. M., Manuelidis, L. & Manuelidis, E. *Nature* **306**, 474-476 (1983).
- Diringer, H. *et al. Nature* **306**, 376-478 (1983).
- Hope, J. *et al. EMBO J.* **5**, 2591-2597 (1986).
- Merz, P. A., Kascak, R. J., Rubenstein, R., Carp, R. I. & Wisniewski, H. M. *J. Virol.* **61**, 42-49 (1987).
- Bolton, D. C., McKinley, M. P. & Prusiner, S. B. *Science* **218**, 1309-1311 (1982).
- McKinley, M. P., Bolton, D. C. & Prusiner, S. B. *Cell* **35**, 57-62 (1983).
- Prusiner, S. B. *et al. Cell* **35**, 349-359 (1983).
- Gibson, P. H., Somerville, R. A., Fraser, H., Foster, J. D. & Kimberlin, R. H. *Vet. Rec.* **120**, 125-127 (1987).
- Dawson, M., Mansley, L. M., Hunter, A. R., Stack, M. J. & Scott, A. C. *Vet. Rec.* **121**, 591-592 (1987).
- Somerville, R. A., Ritchie, L. A. & Gibson, P. H. in *Proc. Symp. Int. sur les Virus Non-conventionnels du Système Nerveux Central* (eds Court, L. A., Dormont, D. & Kingsbury, D.) (Abbaye de Melleray, Moirion la Rivière, in the press).
- McKinley, M. P., Braunfeld, M. B., Bellinger, C. G. & Prusiner, S. B. *J. infect. Dis.* **154**, 110-120 (1985).
- Prusiner, S. B., Groth, D. F., Bolton, D. C., Kent, S. B. & Hood, L. E. *Cell* **38**, 127-134 (1984).
- Oesch, B. *et al. Cell* **40**, 735-746 (1985).
- Meyer, R. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 2310-2314 (1986).
- Manuelidis, L., Valley, S. & Manuelidis, E. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4263-4267 (1985).
- Sklaviadis, T., Manuelidis, L. & Manuelidis, E. E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6146-6150 (1986).

- Hope, J., Multhaup, G., Reekie, L. J. D., Kimberlin, R. H. & Beyreuther, K. *Eur. J. Biochem.* **172**, 271-277 (1988).
- Liao, Y.-C. J., Lebo, R. V., Clawson, G. A. & Smuckler, E. A. *Science* **233**, 346-367 (1986).
- Kretzschmar, H. A. *et al. DNA* **5**, 315-324 (1986).
- Hunter, N., Foster, J. D., Dickinson, A. G. & Hope, J. *Vet. Rec.* (in the press).
- Westaway, D. & Prusiner, S. B. *Nucleic Acids Res.* **14**, 2035-2044 (1986).
- McBride, P. A., Bruce, M. B. & Fraser, H. *Neuropath. appl. Neurobiol.* **14**, 325-336 (1988).
- DeArmond, S. J. *et al. Cell* **41**, 221-235 (1985).
- Dickinson, A. G. & Outram, G. W. in *Novel Infectious Agents and the Central Nervous System Ciba symp.* **135** (eds Bock, G. & Marsh, J.) 63-83 (Wiley, Chichester, 1988).
- Carlson, G. A. *et al. Cell* **46**, 503-511 (1986).
- Carp, R. I., Moretz, R. C., Natelli, M. & Dickinson, A. G. *J. gen. Virol.* **68**, 401-407 (1987).
- Hunter, N., Hope, J., McConnell, I. & Dickinson, A. G. *J. gen. Virol.* **68**, 2711-2716 (1987).
- Hope, J. & Hunter, N. in *Novel Infectious Agents and the Central Nervous System Ciba symp.* **135** (eds Bock, G. & Marsh, J.) 146-163 (Wiley, Chichester, 1988).
- Morgan, K. L. *Vet. Rec.* **122**, 445-446 (1988).
- Fraser, H., McConnell, I., Wells, G. A. H. & Dawson, M. *Vet. Rec.* **123**, 472 (1988).

DNA transformation leads to pilin antigenic variation in *Neisseria gonorrhoeae*

H. Steven Seifert*, Richard S. Ajioka, Christian Marchal*, P. Frederick Sparling* & Magdalene So

Department of Molecular Biology, The Research Institute of Scripps Clinic, La Jolla, California 92037, USA

Many pathogenic bacteria express pili (fimbriae) on their cell surfaces. These structures mediate binding of bacteria to host tissues, and may also be involved in other aspects of pathogenesis. *Neisseria gonorrhoeae* pili are mainly composed of a single protein, pilin, whose expression is controlled at chromosomal expression loci (*pilE*)¹. An intact pilin gene and promoter sequences²⁻⁵ are only found at *pilE*. Strain MS11 contains two expression sites (*pilE1* and *pilE2*)^{1,2}, whereas several of its derivatives⁵ and other clinical isolates⁴ contain only one. Silent pilin loci (*pilS1-pilS7*) contain truncated variant pilin genes lacking the promoter and conserved pilin gene sequences^{3,6}. Pilin antigenic variation in *N. gonorrhoeae* occurs by DNA recombination between one of the silent partial variant gene segments in *pilS* and an expressed pilin gene in *pilE*^{7,8}. The recombination reactions are nonreciprocal, and therefore the mechanism has been classified as gene conversion. We report that much of the recombination between pilin loci actually occurs after transformation of living pilated cells by DNA liberated from lysed cells within a population. This constitutes a new molecular mechanism for an antigenic variation system, as well as the first specific function for a DNA transformation system.

An examination of the biology of the gonococcus suggested that there was an alternative mechanism to gene conversion for pilin antigenic variation. Pilated gonococci are naturally competent for DNA transformation throughout their life cycle⁹, and this process is dependent on homologous recombination^{10,11}.

* Present Addresses: Department of Microbiology and Immunology, Northwestern University Medical School, Chicago, Illinois 60611, USA (H.S.S.); Unite Antigenes Bactériens, Institut Pasteur, 75015 Paris, France (C.M.); Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, North Carolina 27599, USA (P.F.S.).

Fig. 1 Transfer of a pilin marker between strains. *a*, Diagram showing parental bacterial cells and selected progeny. MS11Cm20, was non-piliated (Pil^-) due to a deletion of *pilE1* and a mini-transposon insertion in *pilE2*, chloramphenicol-resistant (Cm^r) due to the mini-transposon, and spectinomycin-sensitive (Spc^s). MS11Spc, was Pil^+ , Cm^s , and Spc^r . The strain that was selected after mixing of the two strains was Pil^+ , Cm^r and Spc^r , having retained the intact *pilE1* locus, acquired the Cm^r gene carried by the mini-transposon in *pilE2* by transformation, and retained the Spc^r allele. *b*, Frequency of Cm^r , Spc^r colonies after various times of mixing. 'DNA' shows the results using purified MS11Cm20 DNA, 'co-cultivation' indicates mixing of the strains, and 'DNaseI', the level of detection of this experiment. No colonies appeared at zero time or after DNaseI treatment.

Methods. A 980-base pair (bp) *Hind*III/*Cla*I DNA fragment encoding pilin was subcloned from pNG1100BH1 (ref. 3) into the shuttle mutagenesis vector pHSS6 (ref. 20), and a mini-transposon insertion isolated that eliminated pilin production in *Escherichia coli* (H. S. Seifert *et al.*, manuscript in preparation). DNA sequencing showed that the mini-transposon insertion was in the DNA corresponding to the first amino acid of the mature protein³. After linearization of the plasmid with restriction endonuclease *Not*I, one non-piliated transformant of MS11 was isolated (MS11Cm20). The transformant reverted to a pilated form at a lower frequency than 10^{-8} per colony forming unit (c.f.u.). The Spc^r marker was introduced into MS11 by transforming with DNA from FA114 (ref. 21). This transformant also received the *penB* marker by co-transformation from FA114 as the level of sensitivity to penicillin or chloramphenicol was representative of FA913 rather than MS11. About 10^8 c.f.u. ml⁻¹ of MS11Spc and MS11Cm20 were treated with 50 μ g ml⁻¹ DNaseI (Sigma, D5025, dissolved in 0.15 M NaCl, 50% glycerol, filter-sterilized, and stored as a 25 mg ml⁻¹ stock solution at -20 °C) or the same enzyme preparation that had been boiled for 5 min. The strains were mixed together in 2 ml liquid GCB (GCB is 15 g proteose Peptone no. 3, (Difco); 1 g corn starch; 4 g potassium phosphate, dibasic; 1 g potassium phosphate, monobasic; and 5 g sodium chloride per litre) with Kellogg supplements²² (GCB+) and 20 mg ml⁻¹ active or inactivated DNaseI. The cultures were grown at 37 °C, 5% CO₂ for 4, 8 or 12 hours, with 20 μ g ml⁻¹ fresh active or inactivated DNaseI added every two hours. The mixed cultures and respective parents were plated on medium containing 75 μ g ml⁻¹ Spc and 1 μ g ml⁻¹ Cm . Colony counts were scored after 40 hours growth, and the presence of a *CAT* gene confirmed by a Southern blot using nick-translated pCM7 (ref. 23) as a probe. The values shown are from a representative experiment. Repeats of the experiment gave similar results.

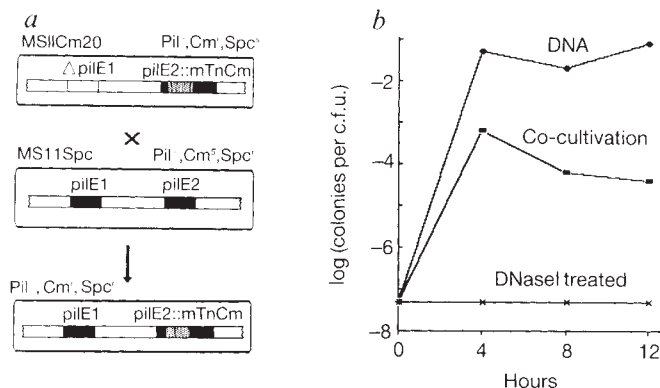
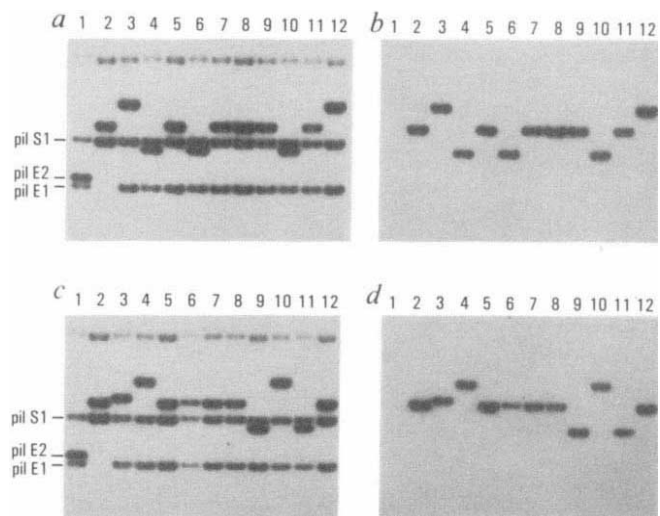


Fig. 2 Southern blot comparing products from mixed culture to free DNA. *a*, *Cla*I digested DNA from MS11 variant A (lane 1), MS11Cm20 (lane 2) and MS11Cm20/MS11Spc mixed-culture isolates (lanes 3-12) probed with nick-translated pilin clone pNG1100. *b*, Duplicate filter to that in *a*, probed with nick-translated *CAT* probe pCM7. *c*, *Cla*I-digested DNA from MS11 variant A (lane 1), MS11Cm20 (lane 2) and transformants of MS11Spc with MS11Cm20 DNA (lanes 3-12) probed with pNG1100. *d*, Duplicate filter to that in *c* probed with pCM7.

Methods. Gonococcal chromosomal DNA was isolated as previously described⁵. Blotting to nitrocellulose filters, and hybridization with nick-translated probes was also as described. Blotting and hybridization to Zeta Probe membrane (Biorad) used the alkaline blotting procedure as recommended by the manufacturer. Oligonucleotide probes were labelled with T4 polynucleotide kinase and purified from unincorporated [γ -³²P] ATP (ICN) by spin dialysis over Sephadex G25 resin (Pharmacia). Nick-translation was as described¹¹ with purification from [α -³²P] dNTPs (Amersham) by spin dialysis over Sepharose CL-6B (Pharmacia).



Although the role of pili in transformation is unclear, pilated variants are transformed at a frequency about 1,000-fold higher than non-piliated variants⁹. *N. gonorrhoeae* naturally undergoes autolysis throughout its growth cycle¹², and DNA released from lysed cells is biologically active for DNA transformation¹³. These observations led us to consider the possibility that antigenic variation occurs by the uptake of silent pilin sequences from lysed cells, and the subsequent recombination of the incoming DNA with partially homologous sequences in the expression site.

If transformation is an important mechanism for antigenic variation, then transfer of pilin sequences between cells in culture by transformation should occur at relatively high frequency. We examined the ability of a marked pilin gene to be transferred between two different gonococcal strains by transformation. A Cm^r , non-piliated transformant of strain MS11 was constructed, where *pilE2* was disrupted by a mini-transposon¹⁴ and the entire

pilE1 locus was deleted (MS11Cm20, Fig. 1a). This non-piliated strain was incompetent for transformation. MS11Cm20 was co-cultured with a pilated, spectinomycin-resistant (Spc^r) MS11 derivative (MS11Spc) in the presence of active or inactivated DNaseI, and the level of Cm^r , Spc^r colonies scored. The *CAT* marker from MS11Cm20 was efficiently transferred to MS11Spc after four hours (Fig. 1b), and the transfer was inhibited by DNaseI only if MS11Cm20 was treated with the enzyme before mixing. A decrease in the number of Cm^r , Spc^r colonies was observed after eight hours with no significant change occurring by 12 hours. Purified chromosomal DNA from MS11Cm20 transformed MS11Spc at high frequency, the appearance of transformants was inhibited by DNaseI, and no increase in the number of transformants was observed with continued exposure of the culture to DNA (Fig. 1b). Progeny obtained from both sets of experiments had the mini-transposon inserted into *pilE2* as shown by Southern blot analysis (Fig. 2). These results indi-

Fig. 3 Construction of gonococcal strain to measure pilin recombination. *a*, Schematic representation of *pilS1*, *pilE1*, and *pilE2* in the MS11 chromosome. Black boxes show 3'-pilin variant sequences, striped boxes indicate 5'-pilin constant sequences, straight lines are non-pilin chromosomal sequences, and wavy lines represent vector sequences. The pilin promoter region is indicated by -10, whereas the ball-on-stick shows the pilin signal sequence. The spacing of the gene segments in the pilin loci are not drawn to scale. *b*, Representation of *E. coli* clone pNG1749CAT. pNG1749 contains a 5-kilobase (kb) *EcoRI* fragment from *pilE2*. *c*, Products of transformation with pNG1749CAT. The *pilE1*CAT and *pilE2*CAT genes are similar and only shown once. The *pilS1*CAT transformant only represents one possible form of this product. *d*, *e*, Southern blot of MS11 transformants. *d*, *ClaI*-digested chromosomal DNA from MS11 variant A (lane 1) and nine Cm^r transformants (lanes 2-10) on Zeta Probe membrane (Biorad) probed with SP3 (ref. 6), an oligonucleotide specific for the *Sma/Cla* repeat of pilin that occurs in every pilin locus. The pilin loci designations are taken from refs 6 and 24. *e*, The same filter as in *d*, stripped of SP3 and reprobed with SP8 (5'-CTTTATAATACGCAGTTGTC), an oligonucleotide that represents a sequence from the pilin promoter region². A separate filter was probed with nick-translated CAT DNA (pCM7) and showed the identical hybridization pattern as *b* (data not shown). **Methods.** pNG1749CAT: the ends of *MstII*-digested pNG1749 and *HindIII* digested pCM7 (ref. 21) were made blunt by filling in with Klenow fragment of DNA polymerase I. The DNA was ligated with T4 DNA ligase and transformed²⁵ into *E. coli* selecting for chloramphenicol- and ampicillin-resistant transformants. One transformant was selected that had the CAT cartridge from pCM7 inserted into the *MstII* site to make pNG1749CAT. All enzymes were purchased from Boehringer Mannheim and used under conditions recommended by the manufacturer. Transformation: gonococcal transformation was by a procedure modified from that of Biswas *et al.*²². Cells were grown on solid GCB+ medium (Difco) for 18 h at 37 °C, 5% CO_2 . About 10^9 cells were collected with a dacron swab into 1 ml GCB liquid medium plus 5 mM MgCl_2 (GCB-Mg) at 37 °C. The cell suspension was vortexed vigorously to disrupt clumps, and diluted 1:10 into GCB-MG with or without DNA (2 mg) at 37 °C. After 30 min at 37 °C without agitation, the cells were diluted 1:10 into GCB+ liquid medium and incubated for 4-6 h at 37 °C in 5% CO_2 in a tissue culture flask (5-7 ml in a 25- cm^2 flask) for expression of phenotypic markers. Cm^r transformants were selected using $10 \mu\text{g ml}^{-1}$ chloramphenicol.

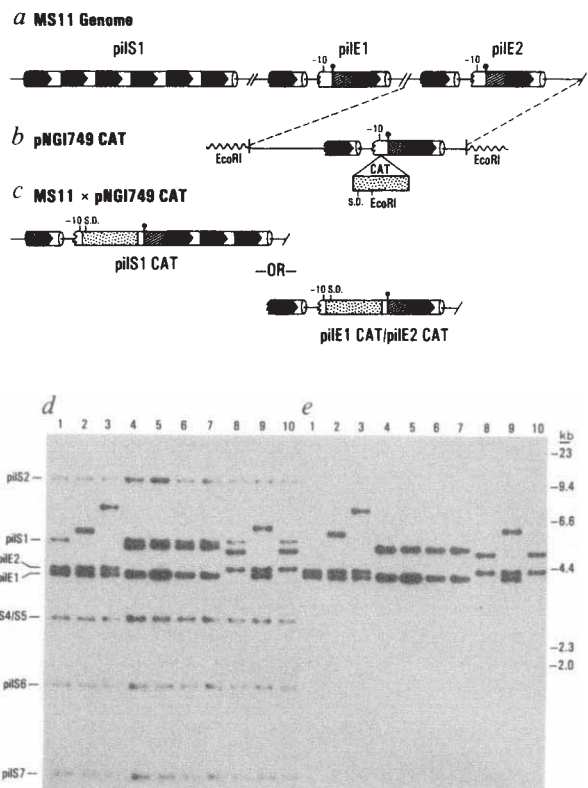
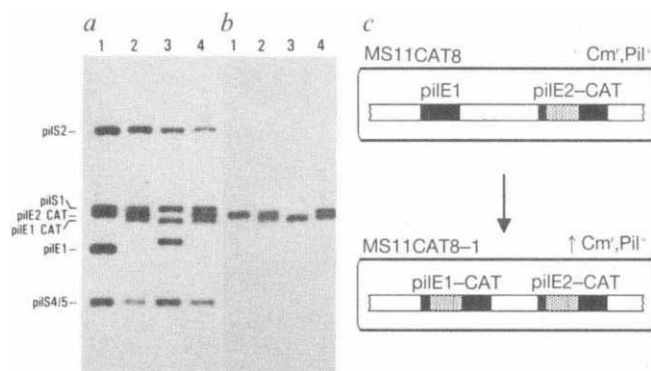


Fig. 4 Model system for assaying non-reciprocal pilin recombination. *a*, *ClaI*-digested chromosomal DNA on nitrocellulose probed with nick-translated pNG1100 DNA¹. *b*, An identical filter probed with nick-translated CAT probe, pCM7. Lane 1, MS11CAT8 (*pilE1*CAT, *pilE2*); lane 2, MS11CAT8-1 (derivative of MS11CAT8 selected on $30 \mu\text{g ml}^{-1}$ Cm); lane 3: MS11CAT5 (*pilE1*, *pilE2*CAT); lane 4, MS11CAT5-1 (derivative of MS11CAT5 selected on $30 \mu\text{g ml}^{-1}$ Cm). *c*, Diagram showing the expression sites *pilE1* and *pilE2* in MS11CAT8 before (top) and after (bottom) selection on high chloramphenicol concentrations.



cate that the Cm^r/Spc^r variants resulted from DNA transformation of pilin-associated chromosomal markers during co-cultivation of the two strains.

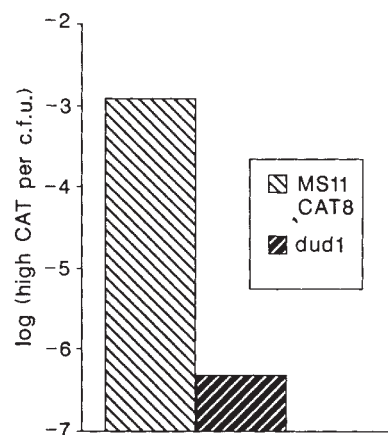
We next examined the nature of pilin-associated recombination events within a single strain. We constructed a gonococcal strain in which CAT gene expression was under pilin promoter control in one *pilE* locus while the other expression locus was unmarked. In constructing this strain, the pilin/CAT fusion was found to have inserted into either expression site as well as *pilS1*, the major silent locus (Fig. 3*d, e*). Insertion into *pilS1* was accompanied by the deletion of various amounts of the *pilS1* sequences (Fig. 3*c*). In these experiments the recombination potential of the silent region that is not transcribed was equivalent to that of the expressed genes that are transcribed, in contrast to reports of eukaryotic systems, where non-transcribed sequences were much less recombinogenic than transcribed genes^{15,16}. Two Cm^r transformants were chosen for further study. MS11CAT8 contains the CAT marker in *pilE1*, whereas MS11CAT5 contains the CAT marker in *pilE2*. Figure 4*a* and *b* shows that in the presence of high levels of chloramphenicol, progeny were selected that contained CAT in both expression sites (this process is shown diagrammatically for MS11CAT8 in Fig. 4*c*). The non-reciprocal transfer of CAT to a second expression locus can be used as a model system to study pilin antigenic variation.

To test whether DNA transformation was involved in the exchange of CAT markers in MS11CAT8, we attempted to inhibit the non-reciprocal exchange of CAT genes with DNaseI. We were unable to show a consistent decrease in the level of appearance of two CAT genes in response to DNaseI treatment.

scribes sequences were much less recombinogenic than transcribed genes^{15,16}. Two Cm^r transformants were chosen for further study. MS11CAT8 contains the CAT marker in *pilE1*, whereas MS11CAT5 contains the CAT marker in *pilE2*. Figure 4*a* and *b* shows that in the presence of high levels of chloramphenicol, progeny were selected that contained CAT in both expression sites (this process is shown diagrammatically for MS11CAT8 in Fig. 4*c*). The non-reciprocal transfer of CAT to a second expression locus can be used as a model system to study pilin antigenic variation.

Fig. 5 Frequency of colonies carrying two *CAT* genes. Bar graph showing the frequency of two *CAT* genes in MS11CAT8 and MS11CAT8*dud1* (*dud1*).

Methods. Transformation of the *dud1* allele from FA660 into MS11CAT8 was performed as described in Fig. 2 except that liquid expression was omitted. Dilutions were made in GCB liquid medium, and plated on GCB solid medium. Cells were grown for 18 h on solid GCB+, collected, and spread on solid medium with $30 \mu\text{g ml}^{-1}$ Cm for MS11CAT8 and $50 \mu\text{g ml}^{-1}$ MS11CAT8*dud1*. The different levels of chloramphenicol used represent the different minimum inhibitory concentration of chloramphenicol for each strain. This was presumably due to the altered colony morphology of the *dud1* strain as both strains produced the same level of *CAT* activity (data not shown). Ten or more variants from three experiments were passaged without chloramphenicol selection and the frequency of variants with two *CAT* genes determined by Southern blots of *Cla*I-digested chromosomal DNA, using nick-translated *CAT* DNA as a probe (pCM7).



As we have shown that pilin DNA is present in gonococcal cultures (Fig. 1), and is active for transformation (Figs 1 and 2), we did not understand why DNaseI had no effect on this process. The data presented in Fig. 1b show that the number of Cm^r , Spc^r cells did not increase with time after the initial co-culturing of the mixed strains. Therefore, if the relative proportion of MS11CAT8 variants with two *CAT* genes was at a maximum before DNaseI treatment, the frequency may be unaffected by DNaseI treatment. In addition, DNA released from autolysed cells may be more accessible to living pilated cells than to DNaseI digestion. This hypothesis is supported by the observation that inhibition of transfer of *pil*⁻ markers between mixed strains by DNaseI (Fig. 1b) could only be demonstrated if the donor strain was treated with the enzyme before mixing with the competent recipient strain. Addition of DNaseI to only the recipient strain before mixing did not reduce the level of Cm^r transferred between strains. These data suggest that pilated cells have a higher affinity for the transforming DNA than DNaseI, and that the enzyme was unable to inactivate the DNA before transformation had occurred.

To examine the role of DNA transformation in pilin antigenic variation without using DNaseI, we assayed the pilin/*CAT* gene conversion events in a DNA-uptake-deficient background. A gonococcal mutant that is deficient in the DNA uptake step of the transformation process but which retains full piliation had been isolated (G. Biswas and P.F.S., manuscript in preparation). This mutation (*dud1*) interferes with the ability of the cell to take up radiolabelled gonococcal DNA into a DNaseI-resistant state, which reduces the transformation frequency of the strain by approximately 1,000-fold. The recombination functions of *dud1* seem to be unaffected as strains carrying this mutation are not sensitive to DNA-damaging agents. Moreover, *dud1* strains can be transformed at a low but measurable frequency compared with wild-type strains, even though DNA uptake—a necessary first step in transformation—is greatly reduced. This indicates that the only effect of *dud1* on transformation is at the level of

DNA uptake, and that homologous recombination is unaffected.

The *dud1* allele was introduced into MS11CAT8 by transformation. The frequency of pilin gene conversion was measured by selection for increased Cm^r , and confirmed by Southern blot analysis. Figure 5 shows that introduction of *dud1* into MS11CAT8 lowers non-reciprocal recombination between pilin genes by about 1,000-fold. The relative frequency of pilin recombination within the *dud1* strain was similar to the relative frequency of transformation found in all *dud1* strains. These data indicate that the majority of non-reciprocal recombination events between pilin expression loci in pilated *N. gonorrhoeae* occur through DNA transformation.

We conclude that the majority of pilin-associated non-reciprocal recombination events observed *in vitro* are the result of DNA transformation, but not that pilin gene conversion *per se* does not occur in the gonococcus. Indeed, as non-piliated gonococci are relatively incompetent for transformation, phase variation (non-piliated to pilated) may occur by an intracellular gene conversion mechanism¹⁷. It has previously been reported that DNA transformation is involved in pilin phase variation^{18,19}. The relative contributions of intracellular and intercellular recombination to phase variation remains to be determined. Certainly in pilated strains the contribution of normal intracellular recombination pathways to pilin antigenic variation seems negligible compared with DNA transformation. This is consistent with the hypothesis that DNA containing double-stranded breaks, such as transforming DNA, is much more recombinogenic than DNA that is stably maintained in the chromosome.

We would like to thank Gour Biswas for supplying *dud1* strain FA660 before publication and R. L. Rodriguez for the *CAT* cartridge. We also thank J. L. Flynn for critical reading of the manuscript and M. Vito for technical help. This work was supported by grants from the National Institutes of Health and National Science Foundation (to M.S.), and from the Damon Runyon-Walter Winchell Cancer Fund (to H.S.S.).

Received 6 June; accepted 14 October 1988.

- Meyer, T. F., Mlawer, N. & So, M. *Cell* **30**, 45–52 (1982).
- Meyer, T. F., Billyard, E., Haas, R., Storzbach, S. & So, M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6110–6114 (1984).
- Haas, R. & Meyer, T. F. *Cell* **44**, 107–115 (1986).
- Swanson, J., Bergstrom, S., Barrera, O., Robbins, K. & Corwin, D. *J. exp. Med.* **162**, 729–744 (1985).
- Segal, E., Billyard, E., So, M., Storzbach, S. & Meyer, T. F. *Cell* **40**, 293–300 (1985).
- Segal, E., Hagblom, P., Seifert, H. S. & So, M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2177–2181 (1986).
- Hagblom, P., Segal, E., Billyard, E. & So, M. *Nature* **315**, 156–158 (1985).
- Swanson, J., Robbins, K., Barrera, O. & Koomey, J. M. *J. exp. Med.* **165**, 1016–1025 (1987).
- Sparling, P. F. *J. Bact.* **92**, 1364–1371 (1966).
- Koomey, J. M. & Falkow, S. *J. Bact.* **169**, 790–795 (1987).
- Koomey, J. M., Gotschlich, E., Robbins, K., Bergstrom, S. & Swanson, J. *Genetics* **117**, 391–398 (1987).

- Hebeler, B. H. & Young, F. E. *J. Bact.* **122**, 355–392 (1975).
- Sarubbi, F. A. Jr & Sparling, P. F. *J. infect. Dis.* **130**, 660–663 (1974).
- Seifert, H. S., So, M. & Heffron, F. in *Genetic Engineering, Principles and Methods* Vol. 8 (eds Hollaender A. & Setlow J.) 123–134 (Plenum, New York, 1986).
- Blackwell, T. K. *et al. Nature* **324**, 585–589 (1986).
- Voelkel-Maiman, K., Keil, R. L. & Roeder, G. S. *Cell* **48**, 1071–1079 (1987).
- Bergstrom, S., Robbins, K., Koomey, J. M., & Swanson, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3890–3894 (1986).
- Norlander, L., Davies, J., Norqvist, A. & Normark, S. *J. Bact.* **138**, 762–769 (1979).
- Baron, E. S. & Saz, A. K. *J. Bact.* **133**, 972–986 (1978).
- Seifert, H. S., Chen, E.-Y., So, M. & Heffron, F. *Proc. natn. Acad. Sci. U.S.A.* **83**, 735–739 (1986).
- Sarubbi, F. A., Blackman, E. & Sparling, P. F. *J. Bact.* **120**, 1284–1292 (1974).
- Biswas, G. D., Sox, T., Backman, E. & Sparling, P. F. *J. Bact.* **129**, 983–992 (1977).
- Close, T. J. & Rodriguez, R. L. *Gene* **20**, 305–316 (1982).
- Hass, R., Schwarz, H. & Meyer, T. F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9079–9083 (1987).
- Dagert, M. & Ehrlich, S. D. *Gene* **6**, 23–28 (1979).

Removal of poly(A) and consequent degradation of *c-fos* mRNA facilitated by 3' AU-rich sequences

Tim Wilson & Richard Treisman*

MRC Laboratory of Molecular Biology, MRC Centre, Hills Road, Cambridge CB2 2QH, UK

The *c-fos* proto-oncogene provides a good system to study the processes underlying messenger RNA degradation. After growth factor stimulation of susceptible cells, the *c-fos* transcription rate transiently increases from a low basal level by as much as 50-fold¹, producing a large amount of exceedingly unstable *c-fos* mRNA that is rapidly degraded¹⁻³. Here, we investigate the *c-fos* mRNA degradation process, and find that: (1) ongoing translation of the *c-fos* mRNA itself is required for its degradation; (2) after synthesis, the mRNA poly(A) tail is rapidly removed, in a translation-dependent manner, leading to accumulation of apparently de-adenylated RNA; (3) deletion or replacement of an AU-rich sequence at the mRNA 3' end significantly stabilizes the mRNA; (4) deletion of the 3' AU-rich sequences dramatically slows the poly(A) shortening rate. These results suggest that the 3' AU-rich sequences act to destabilize the mRNA by directing rapid removal of the mRNA poly(A) tract.

In mouse NIH3T3 cells, *c-fos* mRNA accumulates to high levels by 30 min after serum stimulation, and decays to pre-stimulation levels after 90 min (Fig. 1, lanes 1-6). We used an actinomycin D chase to examine mRNA decay in the absence of continuing transcription. Control experiments showed that addition of actinomycin together with serum is sufficient to block transcription induction completely, demonstrating that the inhibitor acts virtually instantaneously (data not shown). Actinomycin treatment slightly inhibits *c-fos* mRNA turnover (Fig. 1, compare lanes 1-6 with 7-12). When cells are stimulated with serum in the presence of cycloheximide, *c-fos* mRNA is completely stable for at least 90 minutes after actinomycin D addition (Fig. 1, compare lanes 13-15 with 9-12). To test whether cycloheximide acts instantaneously, we added the inhibitor 30 min after serum stimulation and measured RNA levels at later times. Little, if any, degradation was subsequently observed, indicating that cycloheximide acts instantly to block mRNA turnover. Similar results were obtained with *c-fos* mRNA transcribed from the endogenous HeLa cell *c-fos* gene (data not shown). These results are consistent with the notion that turnover of *c-fos* mRNA is sensitive to the inhibition of protein synthesis because it requires ongoing translation of the *c-fos* mRNA itself, rather than continuous synthesis of a labile degradative protein.

Previous work has shown that *c-fos* RNA changes in length before turnover: it has been suggested that this may be due to poly(A) tail shortening^{4,5}. To test this idea, we used a non-denaturing gel assay to analyse HeLa cell *c-fos* transcripts. An RNA probe continuously radiolabeled with ³²P, spanning the mRNA 3' end was annealed to total cell RNA, the resulting duplexes were treated with nuclease T1, and the nuclease-resistant RNA was fractionated by non-denaturing or denaturing gel electrophoresis. Under the conditions used, the poly(A) remains intact and is not cleaved from the mRNA. Upon non-denaturing gel electrophoresis, the mixed single-strand/duplex RNA will migrate with mobility dependent on poly(A) tail length; heterogeneity in poly(A) length will cause migration as a smear rather than a sharp band. The results are shown in Fig. 2. HeLa cells have a somewhat higher basal level of *c-fos* transcription than the 3T3 cells discussed above, but the characteristic transient accumulation of *c-fos* mRNA still occurs after serum stimulation (Fig. 2a, lanes 1-6, lower panel). At early times after

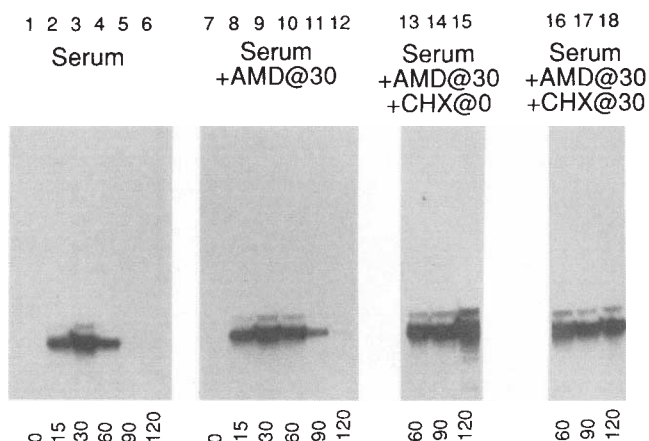


Fig. 1 Translation of *c-fos* mRNA is required for its degradation. Serum-starved NIH3T3 cells were stimulated as indicated. Total cellular RNA was prepared at the times indicated after stimulation and 30 µg analysed for murine *c-fos* mRNA (*c-fos*^M) by ribonuclease protection mapping. Lanes 1-6, RNA from cells at 0, 15, 30, 60, 90 and 120 min after stimulation with 15% fetal calf serum. Lanes 7-12, as lanes 1-6 but with 5 µg ml⁻¹ actinomycin D (AMD) added 30 min after stimulation. Lanes 13-15, RNA at 60, 90 and 120 min after stimulation with serum and 10 µg ml⁻¹ cycloheximide (CHX); 5 µg ml⁻¹ actinomycin D was added 30 min after serum stimulation. Lanes 16-18, RNA at 60, 90 and 120 min after serum stimulation, with both actinomycin D and cycloheximide added 30 min after stimulation.

Methods. Total cell RNA preparation and RNase mapping were as previously described⁴. The probe was generated from plasmid pSP64^{FM}5', which carried an *Aha*11 fragment (nucleotides -14 to +180) spanning the 5' end of the murine *c-fos* gene (*c-fos*^M; ref. 29) inserted in the antisense orientation in the *Eco*RI site of pSP64. Correctly initiated *c-fos*^M RNA generates a 180-nucleotide RNase-resistant fragment from this probe.

stimulation most *c-fos* mRNA molecules carry a long poly(A) tail (Fig. 2a, lanes 2, 3); later, the poly(A) length distribution becomes more heterogeneous (Fig. 2a, compare lanes 2, 3 with 5, 6). At 40-45 min after stimulation, the majority of HeLa cell *c-fos*^H (human *c-fos*) RNA molecules carry only short poly(A) tracts or are de-adenylated (Fig. 2a, lanes 5, 6). Control experiments showed that the smears observed were due to poly(A), as pretreatment of the RNA with RNase H in the presence of oligo(dT) led to the generation of hybrids of the length predicted for those generated by de-adenylated *c-fos*^H RNA (data not shown). To establish a precursor-product relationship between the polyadenylated and de-adenylated *c-fos*^H mRNA, we performed an actinomycin D chase at 15 min after stimulation, a time at which most RNA has a long poly(A) tail (Fig. 2a, lanes 10-16). This experiment confirms that the apparent change in poly(A) length is due to shortening of initially long chains rather than *de novo* synthesis of shorter molecules. In addition, the results suggest that the mRNA is stable until the poly(A) tail has been substantially removed, after which the RNA is degraded (Fig. 2a, lanes 2, 10-16). In the case of *c-myc* RNA degradation *in vitro*, cleavage within an AU-rich region has been reported⁶; however, we did not detect increased cleavage within the *c-fos* AU-rich region during mRNA degradation (data not shown). Rapid poly(A) shortening requires ongoing translation, as serum stimulation in the presence of cycloheximide results in a greatly decreased shortening rate (Fig. 2b, lanes 1-3). Although shortening is observed at later times, this may reflect the toxic effects of the drug.

To examine regions of the *c-fos* mRNA involved in mRNA degradation, we concentrated on the 811-nucleotide long 3' untranslated region. The *c-fos* 3' sequences are required for

* To whom correspondence should be addressed: Imperial Cancer Research Fund, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK.

Fig. 2 Analysis of *c-fos*^H poly(A) shortening rates. RNase T1-resistant hybrids formed between *c-fos*^H RNA and a probe spanning the 3' end of the mRNA were fractionated on non-denaturing (upper panel) or denaturing (lower panel) gels. Times in minutes after stimulation are shown below each lane. *a*, Poly(A) shortening in HeLa cells after serum stimulation. Lanes 1-9, RNA from cells stimulated with serum; lanes 10-16, RNA from serum-treated cells after addition of actinomycin D 15 min after stimulation. *b*, Effect of cycloheximide on poly(A) shortening rate. Lanes 1-3, RNA from cells stimulated with 15% serum and 10 $\mu\text{g ml}^{-1}$ cycloheximide. Arrowheads show positions of pBR322 MspI marker fragments.

Methods. Total cell RNA preparation and RNase protection mapping were performed as described⁴, except that 0.5 $\mu\text{g ml}^{-1}$ RNase T1 only was used for ribonuclease treatments. The 3' probe plasmid, pSP64 F^H3', contains *c-fos*^H nucleotides 3,131 (*Nco*I) to ~4,350 (*Aha*III) inserted in the antisense orientation into pSP64. Probe produced from this plasmid linearized at the *Bcl*I site generates a 240-nucleotide fragment mapping the 3' polyadenylation site. Soon after stimulation large amounts of presumptive *c-fos*^H mRNA precursor are present that generate fragments complementary to the entire probe (data not shown).

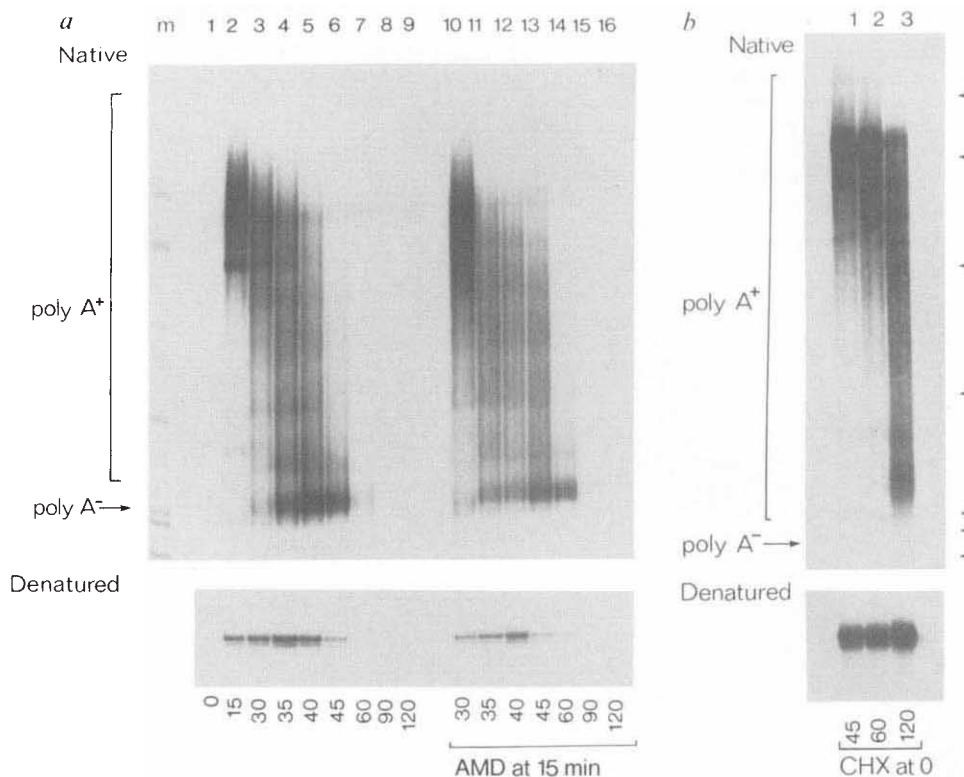


Fig. 3 *a*, Stability of mutant *c-fos*^H mRNAs in transient serum stimulation assays for genes A-E. Total cell RNA from cells transfected with the following *c-fos*^H plasmids was analysed by RNase protection mapping: lanes 1-4, pF711 (ref. 4); lanes 5-8, pF711.BSVA; lanes 9-12, pF711.ΔAU; lanes 13-16, pF711.13; lanes 17-20, pF711.1223. Each set of lanes contains RNA prepared at 0, 40, 90 and 240 min after serum stimulation. Protected fragments generated by the *c-fos*^H and reference α -globin (ref) 5' exons are indicated. *b*, Structures of the genes analysed, *c-fos* 3' AT-rich sequences shown as a solid bar and SV40 sequences cross-hatched; quantitation data, expressed as the ratio of *c-fos*^H RNA present at 90 and 240 min after stimulation relative to the peak value at 40 min, are shown at the right, with standard error of the mean in brackets. The number of independent determinations is shown in brackets at the left.

Methods. Cell transfection, RNA preparation, RNase mapping were as previously described⁴. Plasmids were derived from pF711 (ref. 4) and are as follows. pF711.BSVA has *c-fos*^H (ref. 30) sequences 3' to the *Bcl*I site at nucleotide 3,897 replaced by the early region poly(A) addition signal (*Bam*HI+*Bcl*I fragment, nucleotides 2,771 to 2,534). pF711.ΔAU lacks *c-fos*^H AU sequences (5' 3,949 TGAAAACGTTTTATTGTGTTTTTAATTTATTTATT AAGATGGATTCTCAGATATTTATTTTATTTTATTTTATTTT 4,024 3') and was constructed by site-directed mutagenesis of an appropriate M13 clone using the mutagenic primer 5' CAAAGACC TCAAGGTAGATCTGGAAAACCTGTTAATGTC 3', which inserts the nucleotides GA at the deletion site. pF711.13 lacks *c-fos*^H nucleotides 3,354 (*Nae*I) to 3,897 (*Bcl*I) and contains an insertion of CATG to create an *Nco*I site at the deletion point. pF711.1223 is derived from pF711.13 and contains a duplication of *c-fos*^H nucleotides 3,354 to 3,897. The reference plasmid was π SVH5 α 118 (ref. 4). Data were quantitated by densitometric scanning of suitable autoradiograms using the MRC densitometry facility and *c-fos*^H RNA values normalized with respect to the α -globin reference standard.

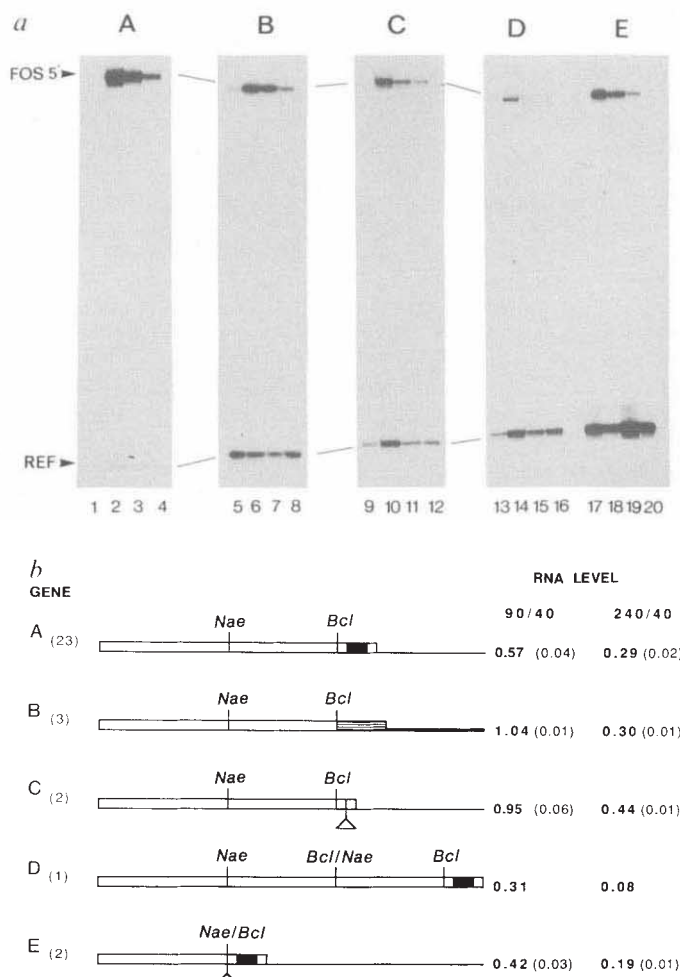
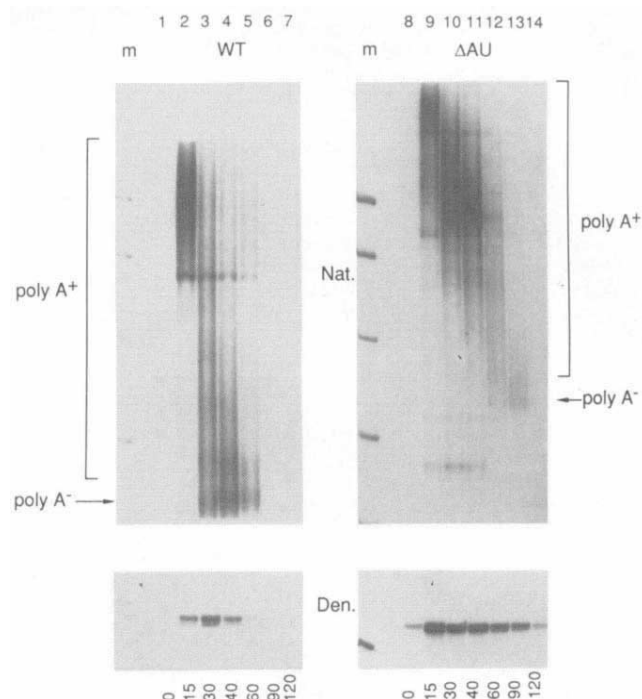


Fig. 4 Analysis of *c-fos*^H 3' poly(A) shortening in transfected NIH3T3 cell pools. RNase T1 resistant hybrids formed between *c-fos*^H RNA and appropriate probes spanning the 3' end of the mRNAs were fractionated on non-denaturing (upper panel) or denaturing (lower panel) gels. Times in minutes after stimulation are shown below each lane. Pools of cells transfected with either pF711 (WT, lanes 1–7) or pF711ΔAU (lanes 8–14) were serum-starved and restimulated with medium containing 15% FCS.

Methods. NIH3T3 cells were transfected with 1 µg supercoiled *c-fos*^H plasmid, 1 µg pSV2neo and 20 µg pUC12 carrier DNA per 75-cm² flask using calcium phosphate co-precipitation as described³¹. Colonies of cells resisting 400 µg ml⁻¹ G418 were pooled and used for RNA analysis. Southern blotting analysis demonstrated that the pools contained insertions of the transfected genes at many different locations (data not shown). A 3' probe plasmid, pSP64F^H3'ΔAU, analogous to pSP64F^H3', but linearized at the *Pvu*II site, was used to analyse transcripts of the mutant gene. De-adenylated mRNA generates protected fragments of lengths 240 nucleotides and 334 nucleotides from wild-type and mutant genes respectively. RNA preparation and mapping were performed as described in the legend to Fig. 2.



transient *c-fos* mRNA accumulation after serum stimulation, at least in part because they destabilize *c-fos* mRNA^{4,7,8}. Previous studies have shown that this region contains a 67-nucleotide AT-rich segment that acts to suppress the transforming activity of the *c-fos* gene^{9,10}; this region is highly conserved and is homologous to sequences in the granulocyte/macrophage colony stimulating factor mRNA that confer instability on rabbit β -globin mRNA¹¹. Studies of histone mRNA degradation have suggested that degradation is dependent on 3' untranslated region length¹². We therefore examined the behaviour of transfected normal and mutant *c-fos*^H genes in a transient serum stimulation assay⁴. Mouse NIH3T3 fibroblasts were transfected with *c-fos*^H gene plasmids together with a plasmid carrying the human α -globin gene as an internal reference standard. Following RNase protection analysis, RNA was quantitated by densitometry and *c-fos*^H RNA levels were normalized to the α -globin reference signal. In the assay, the peak of wild-type *c-fos*^H mRNA accumulation is broader than that of the endogenous gene, reaching a peak at 40 min after stimulation and declining substantially over the next three hours (see Fig. 3, gene A; lanes 1–4; ref. 4). These slower kinetics of mRNA turnover are probably due to the large numbers of template molecules in each transfected cell, as the *c-fos*^H genes in stably transfected mouse NIH3T3 cell lines exhibit similar turnover kinetics to the endogenous mouse gene (see below); nevertheless, mutant phenotypes are clearly discernible in the transient assay. When the extreme 3' end of *c-fos* gene, which contains the AU-rich region, is removed and replaced with the SV40 early region polyadenylation signal mRNA decay is retarded (Fig. 3, gene B; compare lanes 5–8 with 1–4; quantitation in Fig. 3). We used oligonucleotide-directed mutagenesis to construct a gene lacking the AT-rich region (pF711ΔAU; Fig. 3, gene C). This mutant exhibits similar kinetics to the 3' end replacement mutant (Fig. 3, compare lanes 9–12 with 1–4). In contrast, deletion of a 543-nucleotide *Nae*I + *Bcl*II fragment of the 3' untranslated region has essentially no effect on mRNA turnover, although perhaps increasing turnover rate slightly (Fig. 3, gene E; lanes 17–20). Duplication of this region has a similar effect (Fig. 3, gene D; lanes 13–16). These results suggest that the AU-rich sequence at the mRNA 3' end is specifically involved in *c-fos*

mRNA degradation and that proximity of the sequence to the protein coding region is not strictly required for its function. However, it is clear that deletion of the AU-rich sequences does not completely stabilize *c-fos*^H RNA.

We next tested whether the de-adenylation process is affected by deletion of the 3' AU-rich sequences implicated in *c-fos* mRNA destabilization. We generated pools of mouse NIH3T3 cell lines containing either the normal *c-fos*^H gene pF711 or its 3' ΔAU derivative pF711ΔAU by cotransformation of these plasmids with plasmid pSV2neo. In contrast to transiently transfected genes, the stably transfected wild-type *c-fos*^H genes in these cells exhibit kinetics of mRNA induction, poly(A) tail removal and mRNA turnover that are identical to those of the endogenous murine gene (Fig. 4, lanes 1–7, lower panel; compare with Fig. 1, lanes 1–6). In the case of cells stably transfected with the 3' ΔAU mutant gene, poly(A) shortening is both slower and apparently more processive than with the wild-type gene (Fig. 4, compare lanes 8–14 with 1–7); at present we do not know what determines the residual rate of poly(A) shortening seen in this mutant. The mutant RNA is substantially more stable than wild-type *c-fos* mRNA, with appreciable quantities persisting throughout the time course. However, degradation of 3' ΔAU mutant mRNA is still dependent on ongoing translation, as this RNA is stable in the presence of cycloheximide (data not shown).

A number of previous observations have suggested a relation between poly(A) and mRNA degradation. Shortening of poly(A) with age, reminiscent of mRNA 'ticketing'¹³ has been observed in bulk HeLa cell RNA^{14,15} and in several specific cellular RNAs^{16–18}; a correlation between poly(A) length and mRNA stability is observed in some^{19–21} but not all^{22,23} mRNAs. De-adenylated globin mRNA is unstable in microinjected frog oocytes²⁴, and de-adenylated mRNAs can be stabilized by polyadenylation^{12,25}. Our results show that the 3' AU-rich sequences present in shortlived mRNAs, such as *c-fos*^{1–3}, granulocyte/macrophage colony stimulating factor¹¹ and *c-myc*^{6,26}, act to direct rapid shortening of the 3' poly(A) in a translation-dependent manner. The resulting de-adenylated (or oligoadenylated) RNA then forms a labile substrate for further degradation; by analogy with *c-myc* RNA this might occur by endonucleolytic cleavage

at the AU-rich sequences themselves⁶. Because the *c-fos* mRNA is relatively stable until the poly(A) tail is removed, the life of the mRNA is essentially determined by the rate of poly(A) shortening. Interestingly, many of the unstable cellular RNAs induced by growth factor stimulation exhibit such apparently two-stage kinetics of mRNA turnover^{27,28}. It has been proposed that AU-rich sequences destabilize mRNAs by acting as a 'sink' for poly(A) binding proteins⁶. From our data, we prefer the idea that the poly(A) tail and 3' AU-rich sequences are basepaired and that a polysome-associated nuclease cleaves at the mismatched regions. The rapidity with which the *c-fos* poly(A) length becomes heterogeneous suggests that *c-fos* poly(A) shortening may occur by random endonucleolytic cleavage. We speculate that the frequency of mismatches within the basepaired region, the length of the AU-rich region itself, and its distance from the site of polyadenylation will all influence the rate of poly(A) tail shortening and hence the effective life of the mRNA: the *c-fos* 3' Δ AU mutant will be a useful gene with which both to test these ideas and to investigate other *cis* acting determinants of mRNA stability.

We thank Andy Newman and Andrew Smith for advice on *in vitro* mutagenesis and Neo transformation respectively, David Bentley and Hugh Pelham for helpful comments on the manuscript, and Sydney Brenner for arranging a DuPont studentship for T. W. This work was funded by the UK MRC.

Received 30 August; accepted 20 October 1988.

- Greenberg, M. E. & Ziff, E. B. *Nature* **311**, 433-438 (1984).
- Kruijer, W., Cooper, J. A., Hunter, T. & Verma, I. M. *Nature* **312**, 711-716 (1984).
- Muller, R., Bravo, R., Burckhardt, J. & Curran, T. *Nature* **312**, 716-720 (1984).
- Treisman, R. H. *Cell* **42**, 889-902 (1985).
- Mitchell, R. L., Henning-Chubb, C., Huberman, E. & Verma, I. M. *Cell* **45**, 567-574 (1986).
- Brewer, G. & Ross, J. *Molec. cell. Biol.* **8**, 1697-1708 (1988).
- Fort, P. *et al. Nucleic. Acids Res.* **15**, 5567-5667 (1987).
- Rahmsdorf, H.-J. *et al. Nucleic. Acids Res.* **15**, 1643-1649 (1987).
- Miller, A. D., Curran, T. & Verma, I. M. *Cell* **36**, 51-60 (1984).
- Meijlink, F. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4987-4991 (1985).
- Shaw, G. & Kamen, R. I. *Cell* **46**, 659-667 (1986).
- Graves, R. F., Pandey, N. B., Chodchoy, N. & Marzluff, W. F. *Cell* **48**, 615-626 (1987).
- Sussman, M. *Nature* **225**, 1245-1246 (1970).
- Sheiness, D., Puckett, L. & Darnell, J. E. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1077-1081 (1975).
- Sheiness, D. & Darnell, J. E. *Nature* **241**, 265-268 (1973).
- Merkel, C. G., Kwan, S.-P. & Lingrel, J. B. *J. biol. Chem.* **250**, 3725-3728 (1975).
- Medford, R. M., Wydro, R. M., Nguyen, H. T. & Nadal-Ginard, B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5749-5753 (1980).
- Mercer, J. F. B. & Wake, S. A. *Nucleic Acids Res.* **13**, 7929-7943 (1985).
- Wilson, M. C., Sawicki, S. G., White, P. A. & Darnell, J. E. *J. molec. Biol.* **126**, 23-36 (1978).
- Zeevi, M., Nevins, J. & Darnell, J. E. *Molec. cell. Biol.* **7**, 517-525 (1982).
- Paek, I. & Axel, R. *Molec. cell. Biol.* **7**, 1496-1507 (1987).
- Deshpande, A. K., Chatterjee, B. & Roy, A. K. *J. biol. Chem.* **254**, 8937-8942 (1979).
- Muschel, R., Khoury, G. & Reid, L. M. *Molec. cell. Biol.* **6**, 337-341 (1986).
- Marbaix, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **72**, 3065-3067 (1975).
- Huez, G. *et al. Eur. J. Biochem.* **59**, 589-592 (1975).
- Dani, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 7046-7050 (1984).
- Lau, L. F. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1182-1186 (1987).
- Almendral, J. M. *et al. Molec. cell. Biol.* **8**, 2140-2148 (1988).
- Van Beveren, C., van Straaten, F., Curran, T., Muller, R. & Verma, I. M. *Cell* **32**, 1241-1255 (1983).
- Van Straaten, F., van Beveren, C., Curran, T. & Verma, I. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3183-3187 (1983).
- Southern, P. & Berg, P. *J. molec. appl. Genet.* **1**, 327-337 (1982).

Circulation of water within wheat grain revealed by nuclear magnetic resonance micro-imaging

C. F. Jenner*, Y. Xia†, C. D. Eccles†
& P. T. Callaghan†

* Department of Plant Physiology,
Waite Agricultural Research Institute, The University of Adelaide,
Glen Osmond, South Australia, 5064, Australia

† Department of Physics and Biophysics, Massey University,
Palmerston North, New Zealand

Observation of water movement *in situ* is difficult and very few methods are available for measuring directly the motion of water. Movement of solutes or suspended particles is not a reliable guide and the use of radioactive or heavy water is limited to a few applications. Thus, although much has been inferred about the mechanism of water movement through comparatively lengthy pathways¹, less is known for the shorter and more tortuous pathways within and between cells. Still more challenging is the investigation of water motion within structurally complex organs of small dimensions such as the fruits and seeds of plants, some of which behave (for water) as practically closed systems. We have observed longitudinally oriented bulk flow of water in developing grains of wheat using a pulsed gradient spin echo nuclear magnetic resonance technique combined with microscale imaging. Movement of water is associated with import of nutrients by the grain but is on too large a scale to be due to phloem transport alone. The flow observed could be associated with unloading and/or transport of nutrients in the vicinity of the vascular system. This is the first report of the observation *in vivo* of water movement on a sub-millimetre scale by non-invasive methods in any biological system.

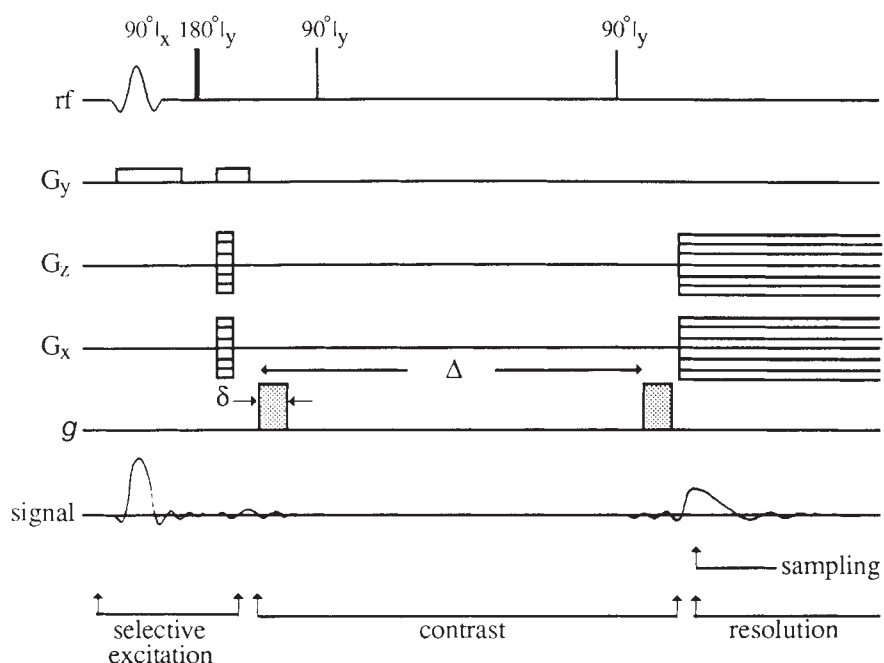
Modern developments in nuclear magnetic resonance (NMR) spectroscopy involving pulsed gradient spin echo (PGSE) techniques combined with imaging methods offer the greatest promise as a means of non-invasively observing the movement of water *in situ* on a millisecond timescale. In NMR imaging, the nuclear spins in a molecule of interest (for example protons

in water molecules) are given a spatial tag by acquiring the NMR precession signal in the presence of a magnetic field gradient. Refinements in methodology² have provided the means of imaging water proton density at a microscopic transverse resolution of around 25 μ m in a 1-mm slice thickness³. PGSE is a well-known variant of the NMR method which involves the formation of a spin echo using two field-gradient pulses separated by a 180° radiofrequency (r.f.) pulse⁴. The position-dependent phase shifts resulting from these pulsed gradients are mutually compensating provided that the nuclear spins remain stationary over the timescale Δ between the gradient pulses. In the event of diffusive motion the pulsed gradients cause echo attenuation, an effect which has been widely exploited in the measurement of molecular self-diffusion⁵. When there is coherent motion, the echo suffers a phase shift, an effect which can in principle be used to measure molecular velocity.

The PGSE and imaging methods may be combined by applying the imaging field gradients to the modulated echo⁵. By independent variation of both the imaging and the PGSE gradients, the spatial positions and the translational motions of nuclear spins can be revealed. The simplest application would involve the observation of attenuation contrast in the image as a result of molecular self-diffusion. A more sophisticated approach uses Fourier analysis of the image modulation as the PGSE gradient pulses are varied in magnitude. Recently it has been shown that this technique enables static spin maps and dynamic nuclear spin displacement profiles to be measured at microscopic resolution in the same experiment⁶. In an earlier study we used simple PGSE contrast methods to detect the regional variation in the diffusion of water in the transverse plane of the developing wheat grain⁷ with a transverse resolution of about 150 μ m. Diffusive motion of water was hindered, especially in the endosperm but less so in regions through which nutrients travel on their way through the grain and into the endosperm.

Transport of nutrients within the grain⁸ involves mass flow of solution acropetally along the vascular bundles of the furrow accompanied by the re-circulation of much, if not all, of the water within the grain. The combined static/dynamic NMR microscopy method⁶ has made it possible to measure simultaneously the motion of water attributable to bulk flow and to

Fig. 1 Radio frequency and gradient pulse sequence used in PGSE imaging experiment. The sections labelled 'selective excitation' and 'resolution' comprise the standard two-dimensional back-projection imaging method. The insertion of the 'contrast' PGSE sequence causes each point in the image to be modulated as a result of molecular motion occurring over the timescale Δ between the PGSE gradient pulses of magnitude g and duration δ . Fourier transformation of the modulation as g is varied provides dynamic displacement profiles from which velocity and diffusion can be calculated at each point in the image. Note that the use of a stimulated echo, in which the usual 180° r.f. pulse is replaced by two widely spaced 90° pulses, permits observation of movement over a much longer timescale than T_2 .



diffusion in a system as small as the wheat grain. We report here the results of preliminary observations indicating that water does indeed flow in certain localized regions of the grain, and that bulk water movement may be associated with the transport of nutrients within it.

Wheat plants (cv 'Otone') were raised in small pots in controlled environment facilities (day/night temperature $27/18^\circ\text{C}$; daylength 12 h) and the ears were selected about 14 days after anthesis when they were approximately half-way through the grain-filling stage of development. The pots were taken to the NMR laboratory immediately before use. Ears were prepared by stripping all except the third or fourth spikelet from the top of the ear and by removing all except the 'b' floret (second from the base) from this spikelet. The rachis was cut immediately above the selected spikelet and the exposed section of rachis below it was wrapped with moist cotton fibre. The pot was inverted and the prepared ear inserted into the NMR probe (4 mm internal diameter) so that the floret was centred in the r.f. coil (5.8 mm internal diameter). The sample compartment was sealed at both ends to prevent evaporation of water. Illumination was provided by a lamp (tungsten halogen) standing 2.3 m from the plants supplying irradiance at about 100 W m^{-2} . For comparison with the attached florets (described above) comparable grains were removed from the ear, thus preventing further import of nutrients, and inserted into the probe.

The NMR microscope used in these experiments is based on a JEOL FX60 NMR spectrometer incorporating gradient coils G_x , G_y and G_z capable of producing 95, 140 and 149 G cm^{-1} respectively. Filtered back-projection reconstruction was used following the stimulated echo pulse sequence shown in Fig. 1. Between the two 90° r.f. pulses the proton spin magnetization is aligned longitudinally, thus protecting it from T_2 relaxation. This sequence permitted the measurement of spin displacement over a timescale of 291 ms despite the short ($\sim 20\text{ ms}$) T_2 relaxation time. Transverse slices were selectively excited using sinc-modulated r.f. pulses. In the dynamic imaging experiment reported here successive real and imaginary (64×64 pixel) images were obtained under successively increased PGSE gradient amplitudes with gradients directed along the grain axis. Acquisition time per image was 1.6 h. Subsequent complex Fourier transformation in the PGSE gradient (q-space) yielded the distribution of dynamic displacements at each pixel in image space⁶.

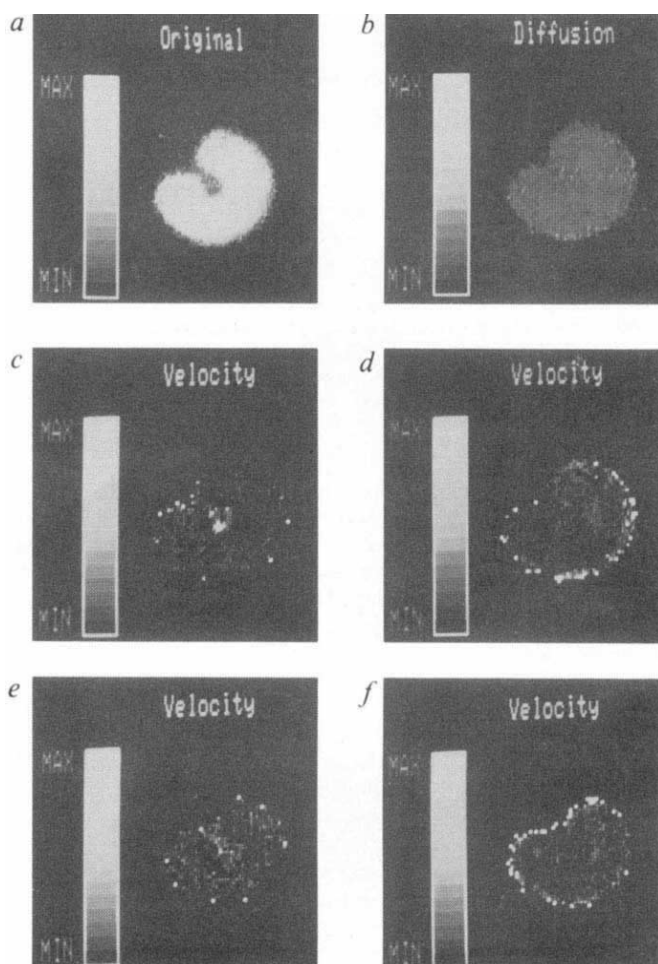


Fig. 2 Image (64×64 pixel) arrays for *a*, water proton density (attached grain); *b*, self-diffusion map (attached grain); *c*, positive acropetal velocity map (attached grain); *d*, negative acropetal velocity map (attached grain); *e*, positive acropetal velocity map (detached grain); *f*, negative acropetal velocity map (detached grain). Note that the water proton density image is significantly T_1 -contrasted because of the stimulated echo sequence used.

a

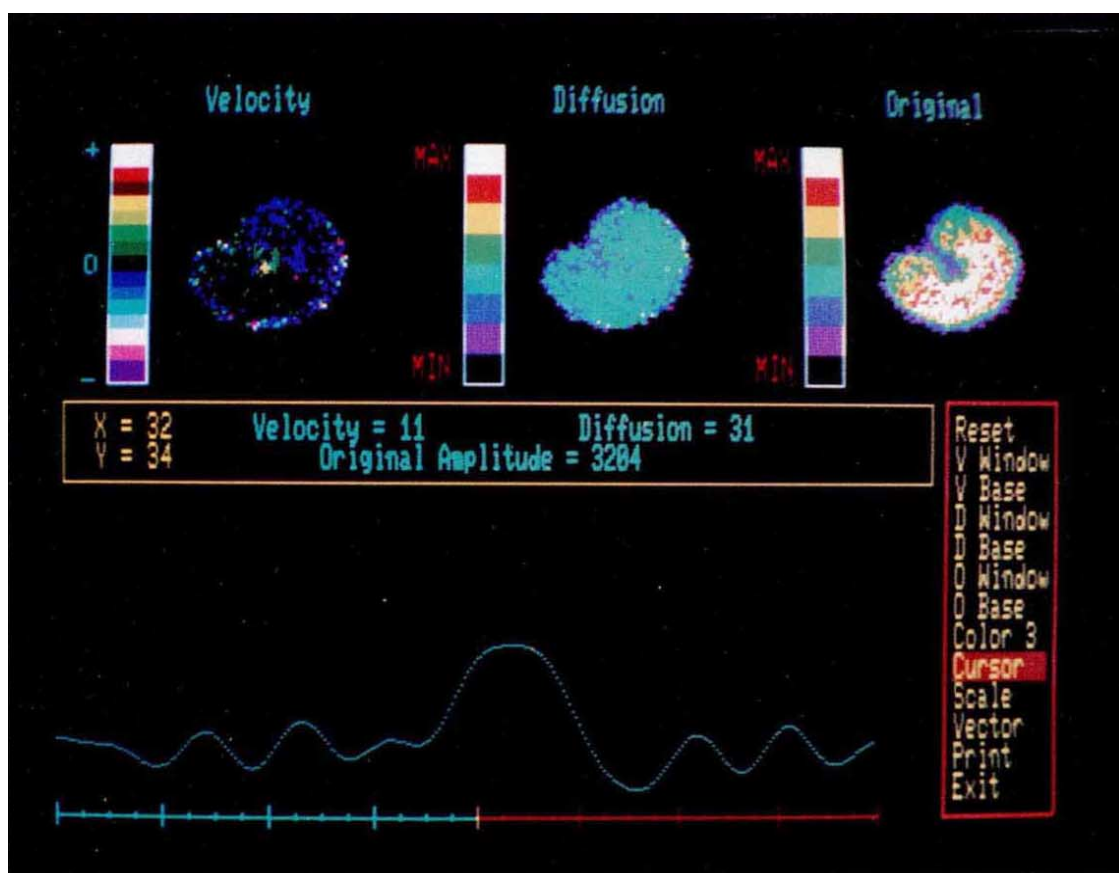
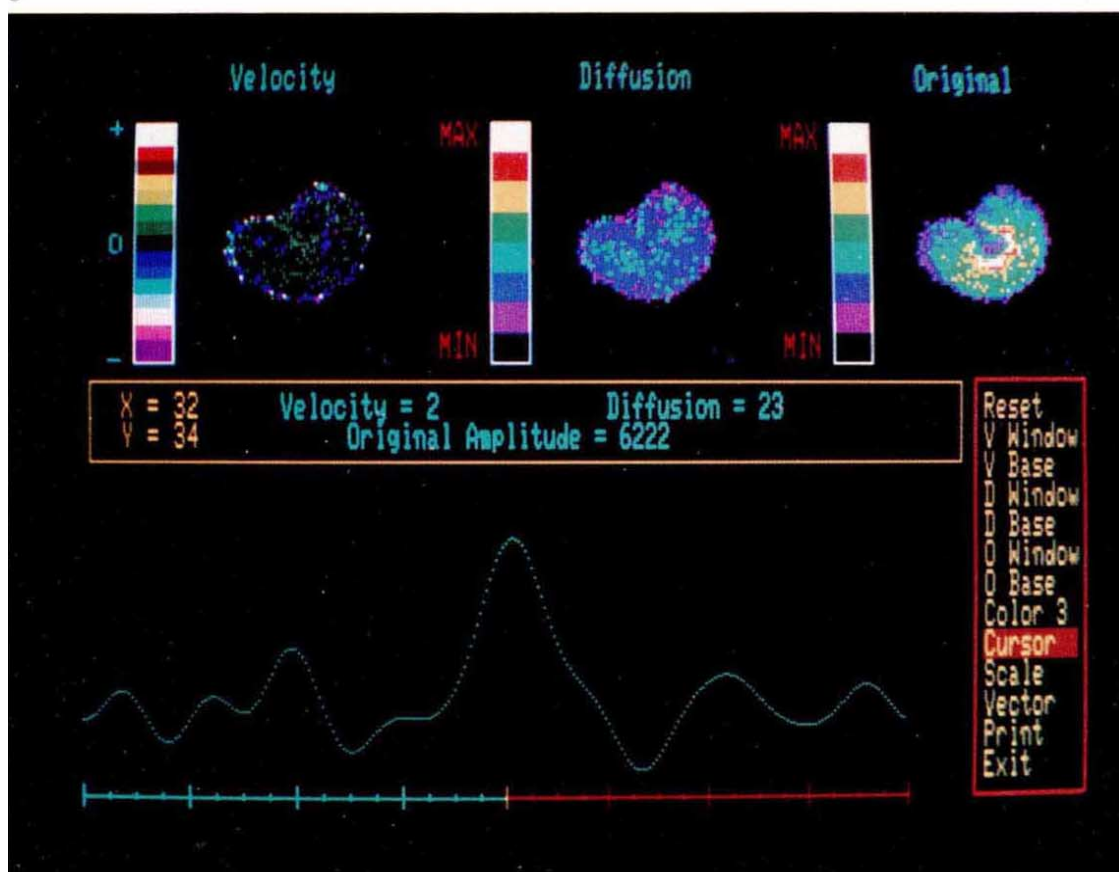


Fig. 3 Velocity, diffusion and original density maps for water in transverse sections of wheat grain, equivalent to the monochrome images of Fig. 2. *a*, Attached grain; *b*, detached grain. In the lower section of each computer graphics display is a displacement profile corresponding to the distribution of velocities in a pixel ($X = 32$, $Y = 34$) where significant positive velocity is exhibited in the *in vivo* grain.

b



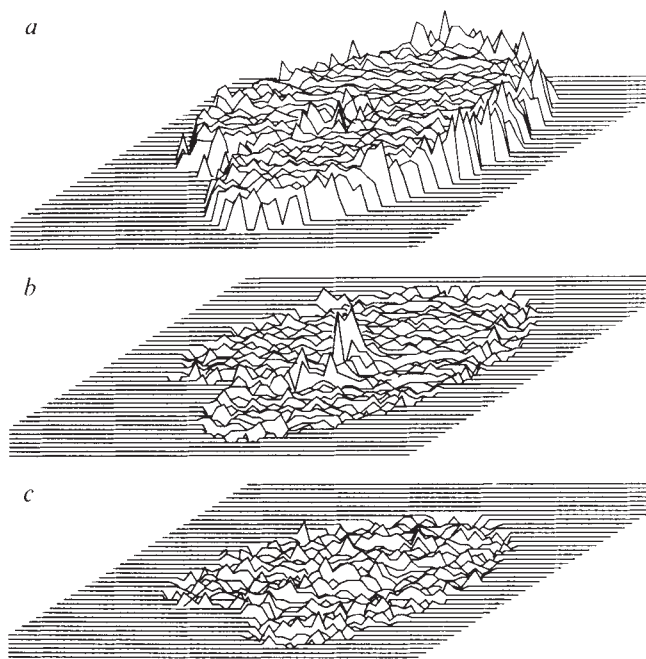


Fig. 4 Stacked profile plots corresponding to the images in Figs 2 and 3. *a*, Self-diffusion *in vivo* map (attached grain); *b*, (signed) acropetal *in vivo* velocity map (attached grain); *c*, (signed) acropetal *in vitro* velocity map (detached grain).

After each dynamic imaging experiment the scanned ear was cut from the culm and cultured⁹ in a solution of [¹⁴C] sucrose for 24 h. After drying (50 °C) the grain was separated from its enveloping bracts. Grains and bracts were digested and their content of [¹⁴C] assayed by liquid scintillation spectrometry.

Several grains in attached florets showed convincing evidence of water movement in the acropetal direction. Figure 2*a* shows the water proton spin density image for an attached grain and the corresponding diffusion map is shown in Fig. 2*b*. The equivalent velocity images are shown in Fig. 2*c* and *d* where the grey scales represent water velocity in the positive and negative directions respectively. Although Fig. 2*c* shows clear evidence of a positive velocity near the centre of the grain, no such movement was detected in the detached grain whose positive and negative velocity images are shown in Fig. 2*e* and *f*.

A few isolated pixels at the perimeter of the images also indicate movement in both directions. Large and spurious fluctuations in apparent velocity can arise from the loss of proton signal at the edge of the floret because the algorithm used to compute the velocity can yield false values in regions where the displacement profiles are noisy. By contrast the central velocity peak arises from several well-defined displacement profiles, an example of which is shown in Fig. 3*a*. Unlike the edge effects, all central velocity peaks are positive. Fig. 4*a* and *b* shows stacked profile plots for attached grain diffusion and velocity; the detached grain velocity profile is shown in Fig. 4*c*. These plots are equivalent to the colour images of Fig. 3 except that

the signed velocity is shown and the spurious edge point velocities have been eliminated *ad hoc*. Eight adjacent pixels (Figs 3*a* and 4*b*) all display values of velocity significantly greater than zero, and in the acropetal direction. This cluster of pixels corresponds to the region of the vascular tissue, nucellar projection and the endosperm cavity in the furrow of the grain (see Fig. 3; ref. 10). The velocity averaged for all eight pixels is $78 \pm 2 \times 10^{-3} \text{ mm s}^{-1}$. The total area of these pixels is about 0.08 mm^2 giving a combined volume flow of $6.2 \times 10^{-3} \text{ mm}^3 \text{ sec}^{-1}$ ($23 \text{ mm}^3 \text{ h}^{-1}$). Signal-to-noise ratios for these velocity values were high, as were those for the corresponding attached grain diffusion coefficients, $8.3 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ (Figs 2*b*, 3*a* and 4*a*). The pixel dimension is $100 \mu\text{m}$. Note that the velocity resolution is two orders of magnitude greater than that used previously⁶.

Scattered, apparently at random, throughout the region corresponding to the endosperm, some single pixels in Fig. 3*a*, or small clusters of them, gave low negative values for velocity indicative of basipetal flow. Although it is unlikely that these are due to chance, reliable evaluation of them is precluded because of the very delicate nature of the image phase shifts from which they arise.

In all but one of the attached florets scanned in the NMR probe, ¹⁴C accumulated in the bracts and the grain of the ears cultured after the scanning experiment indicating that these grains were still capable of importing nutrients. One floret showed no signs of ¹⁴C import, and no water movement was evident either. Also the pericarp of this grain seemed senescent, unlike the other grains with bright green pericarps. Its water content of 51% of fresh mass was lower than values typical for other grains (62–67%) and dry mass (29 mg) was higher than the others (17–20 mg). Failure of this particular grain to import nutrients was therefore possibly due to its being at a more advanced stage of development than the others, where grain-filling had ceased.

As mass-flow in the sieve tubes at the mid-point of the grain accounts for volume flow of about $0.06 \times 10^{-3} \text{ mm}^3 \text{ s}^{-1}$ the observed flow exceeds that attributable to phloem transport by at least two orders of magnitude. Movement of water on a substantial scale must therefore be taking place in the vicinity of the furrow of the grain. If motion is not oriented strictly parallel to the long axis of the grain (so far there is no information on this question), then the observed volume flow underestimates the actual rates of water movement. As the wheat grain behaves like an almost closed system¹¹ and loss of water from the grain during scanning was prevented, acropetal flow of water must have been matched by an equivalent flow in the opposite (basipetal) direction. It is therefore tempting to suggest that the negative values noticed in the endosperm are indicative of water returning slowly through a comparatively large area of the grain. Although not on the scale observed here, re-circulation of water within the grain has been predicted on other grounds⁸. As water movement seems to be correlated with import and transport of assimilates into and within the grain, the possibility that the observed flow of water is associated with unloading and/or distribution of assimilates in the vicinity of the furrow justifies further investigation. Observations on flow in the transverse direction and the relationship between the pattern of movement and the structure and development of the grain are planned.

We thank I. Broonig, Department of Scientific and Industrial Research, Palmerston North for the wheat plants.

Received 18 August; accepted 3 October 1988.

- Boyer, J. S. *A. Rev. Pl. Physiol.* **36**, 473–516 (1985).
- Mansfield, P. & Grannell, P. K. *Phys. Rev.* **B12**, 3618–3134 (1975).
- Eccles, C. D. & Callaghan, P. T. *J. magn. Reson.* **68**, 393–398 (1986).
- Stejskal, E. O. & Tanner, J. E. *J. chem. Phys.* **42**, 288–292 (1965).
- Callaghan, P. T. *Aust. J. Phys.* **37**, 359–387 (1984).

- Callaghan, P. T., Eccles, C. D. & Xia, Y. *J. Phys. E* **21**, 820–822 (1988).
- Eccles, C. D., Callaghan, P. T. & Jenner, C. F. *Biophys. J.* **53**, 77–81 (1988).
- Jenner, C. F. in *Phloem Transport* (eds Cronshaw, J., Lucas, W. J. & Giaquinta, R. T.) 271–281 (Liss, New York, 1986).
- Jenner, C. F. *Aust. J. Pl. Physiol.* **3**, 337–347 (1976).
- Thorne, J. H. *A. Rev. Pl. Physiol.* **36**, 317–343 (1985).
- Jenner, C. F. *Aust. J. Pl. Physiol.* **9**, 69–82 (1982).

Structure and assembly of protocatechuate 3,4-dioxygenase

D. H. Ohlendorf*, J. D. Lipscomb† & P. C. Weber*

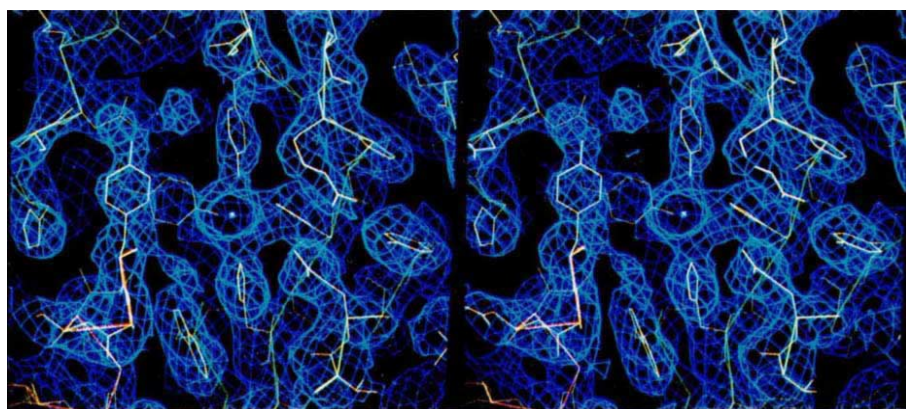
* E. I. du Pont de Nemours and Co., Central Research and Development Department, Experimental Station, PO Box 80228, Wilmington, Delaware 19880-0228, USA

† Department of Biochemistry, 4-225 Millard Hall, University of Minnesota, Minneapolis, Minnesota 55455, USA

Dioxygenases catalyse the cleavage of molecular oxygen with subsequent incorporation of both oxygen atoms into organic substrates. Some of the best-studied dioxygenases have been isolated from bacteria where they catalyse the critical ring-opening step in the biodegradation of aromatic compounds. These bacterial enzymes generally contain nonheme ferric iron as the sole cofactor. Protocatechuate 3,4-dioxygenase (3,4-PCD) was one of the first such enzymes recognized¹ and catalyses the intradiol cleavage of protocatechuic acid by oxygen to produce β -carboxy-*cis,cis*-muconic acid. Previous studies have shown that the 3,4-PCD found in *Pseudomonas aeruginosa* is an oligomer with a relative molecular mass (M_r) of 587,000 (587K) containing 12 copies each of α (22.3K) and β (26.6K) subunits²⁻⁶. The X-ray structure determination of 3,4-PCD reveals the catalytic iron environment required for oxygenolytic cleavage of aromatic rings and also provides a novel holoenzyme assembly with cubic 23(*T*) symmetry and first examples of mixed β -barrel domains.

The structure of 3,4-PCD was solved using a combination of isomorphous replacement and molecular averaging to produce the map in Fig. 1. This map shows continuous electron density for both chains and accounts for all but the first three and last two residues of the β chain. The course of the polypeptide chain was defined initially by α -carbon positions 3.8 Å apart using the graphics program FRODO (ref. 7). An atomic model was constructed from the trace using fragment superposition^{8,9} from a database of refined protein structures selected from the Brookhaven Protein Data Bank¹⁰. Coordinates of fragments in the database corresponding most closely to the observed C_α backbone were inspected, and those best fitting the electron density were incorporated into the X-ray model. Segments of regular secondary structure are well represented in the database and many matches were found for the α -helical and β -sheet regions. Side chains were likewise added to the polypeptide backbone by displaying the low-energy rotamers for each residue¹¹ and retaining that best fitting the observed electron density. After some final manual adjustments, the model was regularized using the REFI feature of FRODO (ref. 7).

Fig. 1 Electron density map at 2.8 Å resolution around the iron at the active site. The α -chain C_α s are connected by red lines, β -chain C_α s by green lines. Electron density extending from the iron atom towards the viewer corresponds to an equatorially bound solvent molecule. The electron density was calculated using native data and data from two heavy atom derivatives, $\text{Pt}(\text{CN})_6^{2-}$ and tetrakis(acetoxymecuri)-methane, collected on a Nicolet Imaging Proportional Counter and processed using the XENGEN data-reduction software²². Previous knowledge of the location of the molecular symmetry axes² was used to find the main heavy atom sites using the minimum function^{23,24}. Additional sites were found using cross Fourier difference maps, refined²⁵ and included in the calculation of the initial 5 Å and 2.8 Å maps. The former map was used to determine a molecular boundary which was used in conjunction with the molecular symmetry to implement the algorithms of Brice²⁶ and Wang²⁷ to iteratively improve the 2.8 Å map. During this process the phases shifted an average of 73° and the correlation coefficient between calculated and observed structure factors rose from 0.47 to 0.82.



The active-site iron associated with each protomer was identified as a 12 σ peak in a symmetry-averaged anomalous difference Fourier map at 2.8 Å resolution. The crystallographic *R*-factor for 59,057 calculated and observed structure factors from 10 to 3.0 Å resolution was 0.38 for this initial model. Although X-ray refinement will reveal detailed differences among crystallographically independent protomers, the symmetry-averaged map provides a clear overall view of the molecular architecture.

The α and β subunits of 3,4-PCD, respectively incorporating 200 and 238 amino-acid residues, share a common structural core but differ in their termini and in several surface loops (Fig. 2). The central structural feature in both subunits is composed of a flattened 9-stranded β barrel and three short α helices (Fig. 4a). The flattened β barrels of 3,4-PCD consists of an antiparallel sheet (strands S3, S4, S6 and S7) and a mixed sheet (strands S2, S5 and S9) which are connected by the shorter strands S1 and S8. The small helical domain is made up of three short α helices connecting S7 and S8 and serves to cover part of one side of the barrel through interactions with strands S7 and S9. The 3,4-PCD barrels are unusual because they incorporate mixed β sheet; the β barrels of other proteins consist either exclusively of parallel or antiparallel chains^{12,13}.

Although the β subunit is 38 residues longer than the α subunit, comparison shows that the common secondary structures include some 129 amino-acid residues whose C_α coordinates superimpose with an r.m.s. error of 1.3 Å. In fact, the subunit core structures within a protomer are related by pseudo-2-fold symmetry (169° rotation and 6.8 Å translation). The subunit interactions involve in part the antiparallel sheet strands S2-S3, S4-S5, S6-S7 and their connecting loops, which are relatively similar in structure and sequence between subunits (Fig. 4b). Subunit differences are much larger in other parts of the structure, reflecting the quite different interactions made by the subunits of the protomers in forming the holoenzyme.

The 12 $\alpha\beta$ protomers of 3,4-PCD are assembled to form a rough truncated tetrahedron with 180 Å edges. Protomers pack together at 2-fold axes that connect centres of opposite edges of the tetrahedron, and at 3-fold axes passing through the particle along lines that connect the centres of faces and their opposite apices. Nearly all of the protomeric contacts are provided by the β subunits that form a hollow inner shell bounding a central cavity of about 50 Å in diameter. The α subunits are clustered around the 3-fold axes and occupy the apices of the tetrahedral aggregate.

Protomeric interactions around the local 2-fold symmetry axes are intimate and involve multiple loop-loop contacts (Fig. 3a). The interacting regions correspond to sequence insertions into loops connecting core β -sheet strands S1-S2, S3-S4, S5-S6 and

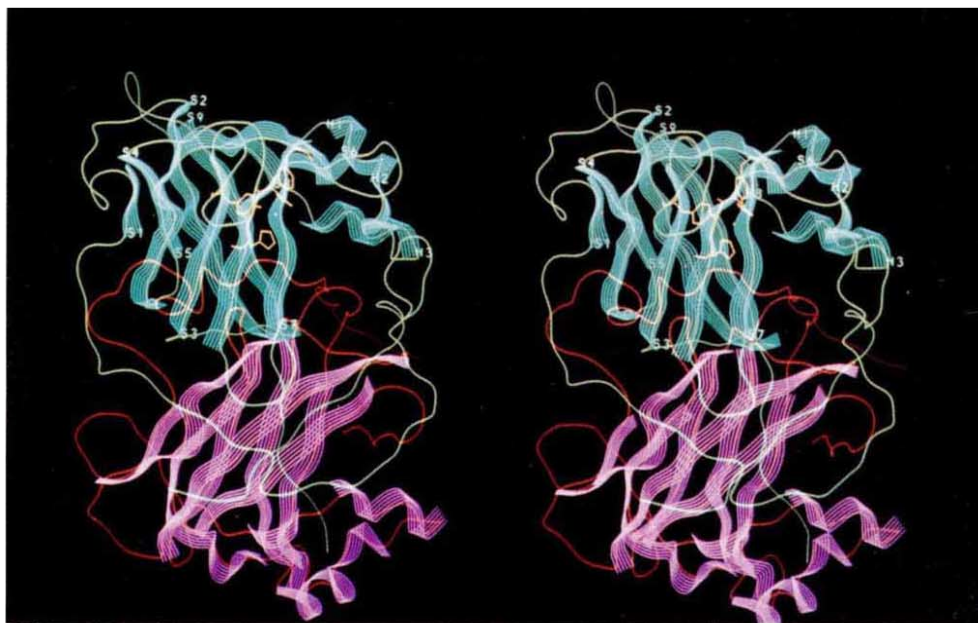


Fig. 2 Stereoview of the 3,4-PCD protomer main chain. The α subunit is red and the β subunit is green. The α subunit secondary structural features (labelled at amino termini) are shown as magenta ribbons and those of the β subunit as cyan ribbons. The iron atom and its ligands are gold.

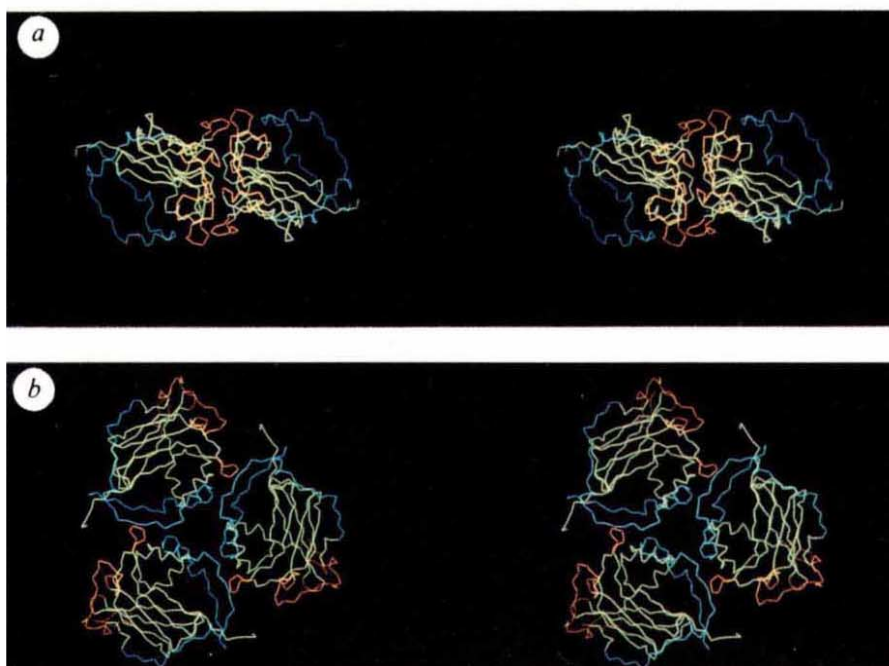


Fig. 3 C_{α} stereo drawing of β -subunit interactions around symmetry axes of the holoenzyme. *a*, 2-fold protomer packing; *b*, 3-fold interactions. The amino-terminal 55 residues are cyan, and connecting loops between strands S1-S2, S3-S4, S5-S6 and S8-S9 are gold.

S8-S9 in the β subunits (Fig. 4*b*). The most extensive 3-fold interaction in the protomer involves mainly β subunit residues in the amino-terminal extension and the loop connecting S5-S6 in the subunit core (Fig. 3*b*). Although the α subunits pack around the 3-fold axis where it emerges from an apex of the holoenzyme, they do not make extensive contact, with the result that channels of 9 Å diameter remain open to the interior of the particle. At the positions where the 3-fold axes emerge from the faces of the tetrahedral aggregate, even larger channels of 16 Å diameter are present.

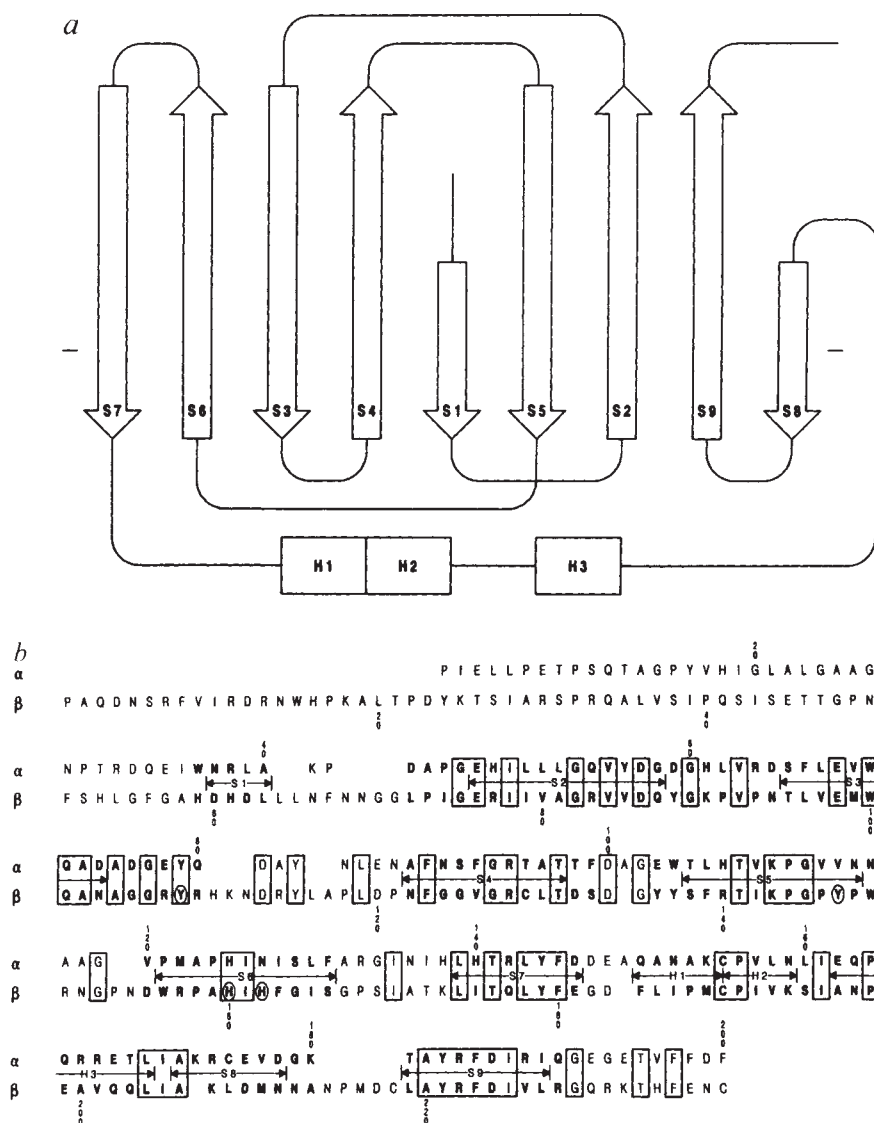
The active site of the enzyme is located at the interface between the α and β subunits. The catalytic ferric ion is coordinated by four protein ligands, Tyr 118 β , Tyr 147 β , His 160 β and His 162 β , which are on sheet strands that extend from the β subunit barrel and curl over the α subunit (Fig. 2). The iron coordination geometry forms an approximate trigonal bipyramid with Tyr 147 β and His 162 β located in axial coordinating positions, and a bound solvent molecule completing the trigonal iron coordination in the equatorial plane (Fig. 1). This coordination is geometrically similar to iron superoxide

dismutase^{14,15} and is consistent with proposals based on spectroscopic and EXAFS measurements^{16,17}.

The ferric ion lies at the bottom of a flat, binding pocket bounded by its ligands, Tyr 24 β , Trp 149 β , Arg 157 β , Ile 191 β and α -chain segments 12-16 and 133-135. Modelling studies of bound aromatic substrates suggest that the guanidino group of Arg 133 α , located at the entrance of the pocket, could serve as a counterion for the protocatechuate carboxylate group when bound in the active site. X-ray studies of enzyme-inhibitor complexes are in progress to provide more details about substrate binding and a better understanding of the catalytic mechanism of the 3,4-PCD enzyme.

Both the amino-acid sequence homology (33%) and structural similarity of the α - and β -subunit cores suggest that the protomer may have diverged from an ancestral homodimer. Examination of the aligned sequences (Fig. 4*b*) shows that Tyr 79 α , Val 114 α , His 125 α and Asn 127 α occupy positions corresponding to the observed iron ligands Tyr 118 β , Tyr 147 β , His 160 β and His 162 β . Thus, two out of four β -subunit iron ligands are present in the α subunit. However, as we have already noted,

Fig. 4 *a*, Schematic diagram of the conserved subunit fold in 3,4-PCD. The β strands (S1-S9) are shown as arrows and α helices (H1-H3) as boxes. *b*, Sequences of α and β chains aligned using observed structural similarities. Corresponding residues are bold, identities boxed, iron ligands circled, and locations of β strands (S1-S9) and α helices (H1-H3) indicated. No attempt has been made to align the sequences of the amino termini. Features of this alignment have been recognized previously⁶.



interactions from the amino terminus of the α subunit are also involved in forming the active site in the β subunit. The corresponding amino-terminal segment of the β subunit in the protomer is repositioned so that Ser 44 β hydrogen bonds to Tyr 79 α and His 125 α in lieu of a ferric ion. It is unclear whether this arrangement developed independently or is derived from a symmetric ancestral dimer with two iron-binding sites. What is evident is that the amino termini of the α and β subunits have functionally and structurally diverged, so that the former has a role in substrate binding at the active site (Fig. 1) and the latter forms the main contacts for protomer assembly around the particle 3-fold axes (Fig. 3b).

It is interesting to note that although all 3,4-PCDs studied so far share a common $\alpha\beta\text{Fe}^{3+}$ protomer organization, there is significant diversity in the quaternary structure of the holoenzyme¹⁸⁻²⁰. At present the functional significance of the holoenzyme organization is unclear as there is no evidence for subunit cooperativity in 3,4-PCD. Continuing structural studies on other oligomeric species of 3,4-PCD, including an $(\alpha\beta)_4$ form in *Pseudomonas cepacia*^{18,21} and $(\alpha\beta)_5$ in *Brevibacterium fuscum*¹⁹, should provide insight into how subunit diversification leads to novel holoenzyme assemblies.

We thank F. R. Salemme for discussion. This work was supported in part by NIGMS (J.D.L.). C_α coordinates will be submitted to the Protein Data Bank.

Received 29 July; accepted 7 October 1988.

- Stanier, R. Y. & Ingraham, J. L. *J. biol. Chem.* **210**, 799-808 (1954).
- Ohlendorf, D. H., Weber, P. C. & Lipscomb, J. D. *J. molec. Biol.* **195**, 225-227 (1987).
- Kohlmeier, N. & Howard, J. B. *J. biol. Chem.* **254**, 7292-7308 (1979).
- Kohlmeier, N. & Howard, J. B. *J. biol. Chem.* **254**, 7309-7315 (1979).
- Iwaki, M., Kagamiyama, H. & Nozaki, M. *J. Biochem.* **86**, 1159-1162 (1979).
- Iwaki, M., Kagamiyama, H. & Nozaki, M. *Archs Biochem. Biophys.* **210**, 210-223 (1981).
- Jones, T. A. *J. appl. Crystallogr.* **11**, 268-272 (1978).
- Jones, T. A. & Thirup, S. *EMBO J.* **5**, 819-822 (1986).
- Finzel, B. C. *Proteins: Structure, Function and Genetics* (in the press).
- Bernstein, F. C. *et al. J. molec. Biol.* **112**, 535-542 (1977).
- Ponder, J. W. & Richards, F. M. *J. molec. Biol.* **195**, 775-791 (1987).
- Richardson, J. S. *Nature* **268**, 495-500 (1977).
- Rossmann, M. G. & Argos, P. *Rev. Biochem.* **50**, 497-532 (1981).
- Ringe, D., Petsko, G. A., Yakamura, F., Suzuki, K. & Ohmori, D. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3879-3883 (1983).
- Stallings, W. C., Powers, T. B., Patridge, G. A., Fee, J. A. & Ludwig, M. L. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3884-3888 (1983).
- Pyrz, J. W., Roe, A. L., Stern, L. J. & Que, L. Jr *J. Am. chem. Soc.* **107**, 614-620 (1985).
- Whittaker, J. W. & Lipscomb, J. D. *J. biol. Chem.* **259**, 4487-4495 (1984).
- Bull, C. & Ballou, D. P. *J. biol. Chem.* **256**, 12673-12680 (1981).
- Whittaker, J. W., Lipscomb, J. D., Kent, T. A. & Munck, E. *J. biol. Chem.* **259**, 4466-4475 (1984).
- Durham, D. R., Stirling, L. A., Ornston, L. N. & Perry, J. J. *Biochemistry* **19**, 149-155 (1980).
- Ludwig, M. L., Weber, L. D. & Ballou, D. P. *J. biol. Chem.* **259**, 14840-14842 (1984).
- Howard, A. J. *et al. J. appl. Crystallogr.* **20**, 383-387 (1987).
- Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. *Acta crystallogr.* **A43**, 1-5 (1987).
- Argos, P. & Rossmann, M. G. *Acta crystallogr.* **B22**, 2975-2979 (1976).
- Terwilliger, T. C. & Eisenberg, D. *Acta crystallogr.* **A39**, 813-817 (1983).
- Bricogne, G. *Acta crystallogr.* **A32**, 833-847 (1976).
- Wang, B. C. *Meth. Enzym.* **115**, 90-112 (1985).

GUIDE TO AUTHORS

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully.

Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

CATEGORIES OF PAPER

Review Articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50–80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

News and Views articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

Scientific Correspondence is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

Submission. All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

GENERAL

All contributions submitted for publication in *Nature* should conform with these rules.

Manuscripts should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

Titles should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

Colour Artwork. A charge of £500 a page is made for 4-colour figures. Inability to meet these costs will not prevent the publication of essential colour figures if the circumstances are explained.

Figure legends must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

References should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unrefereed abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963–65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

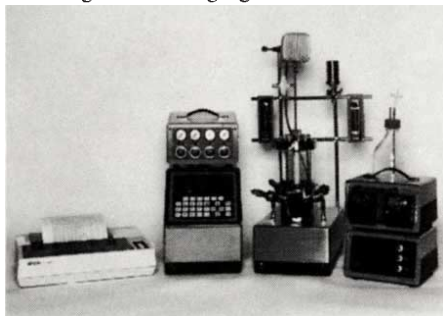
Footnotes should not be used except for changed addresses or to identify the corresponding author if different from the first-named.

Acknowledgements should be brief. Grant numbers and contribution numbers are not allowed.

Scientific exports from Switzerland

As an accompaniment to this week's feature on Science in Switzerland, some of the best and most innovative Swiss products are outlined below, from snake venoms to an industrial ball mill.

CHEMAP AG has scaled down its laboratory and pilot-scale fermenters to come up with a **mini-fermenter** that is just the right size for small production needs — the Chemap CMF (*Reader Service No. 101*). Chemap's mini-fermenter has a working volume of only 750 ml, but it has the same design criteria as larger Chemap fermenters, facilitating later scale-up. The fermenter vessel can be adapted for use with organisms ranging from microbes to



The CMF fermenter with ChemCell.

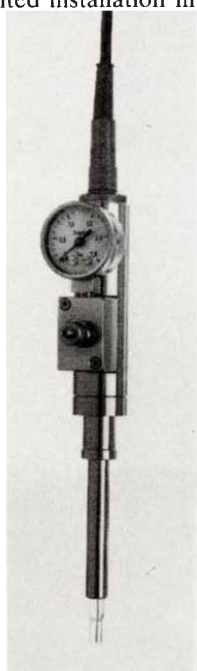
eukaryotic cells, by adding Chemap's ChemCell vibrating mesh unit for cell culture. The vessel can be sterilized *in situ*, and the unit is bottom-stirred through a magnetic coupling. The sensor ring is located at the bottom of the vessel to reduce working volume, and the bottom portion of the fermenter is water-jacketed to allow even heat transfer without localized overheating. Chemap's CMF is available with a wide selection of accessories, including gas blending attachments, mass flow meters and load cells.

Molecular Design has added catalogues from three new manufacturers to its **structural database of chemicals** (*Reader Service No. 102*). The company's Fine Chemicals Directory — available as an in-house program or through the ChemQuest on-line database — now includes compounds from Echo Resins and Laboratory, IndoFine Chemical Co. and JBL Scientific, bringing the total number of catalogues included in the directory to 61. Echo Resins and Laboratory specializes in optical isomers, high-purity glycidyl ethers, photoinitiators and synergists, and functional monomers. IndoFine Chemical Co. produces coumarins, flavones, flavanones and isoflavones, in both research and bulk quantities. JBL Scientific performs custom syntheses, and also supplies high-purity substrates, buffers and proprietary products for clinical diagnostics. The newly released ChemQuest and Fine Chemicals Directory updates contain information on 209,708 items,

representing 65,190 chemicals with a unique structure.

Quantitative aids

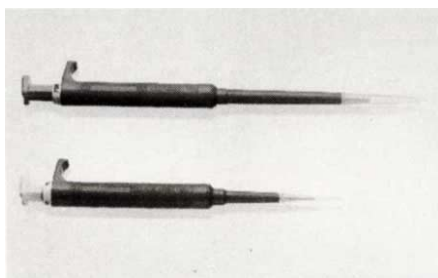
The Infit 767 from Ingold is a **combined pH electrode** and pressurized housing designed for top-mounted installation in bench-scale and small bioreactors (*Reader Service No. 103*). Ingold says both the electrode and the housing of the Infit 767 are autoclavable, to protect against bacterial, fungal or viral contamination of the culture. The sterility of the reference electrolyte is maintained after autoclaving by the presence of an integral microporous filter on the inlet pressure line. The pH electrode has been designed to be dismantled without the need for special tools, and has screw-capped electrodes to simplify cable connection and the exchange



The robust Infit 767.

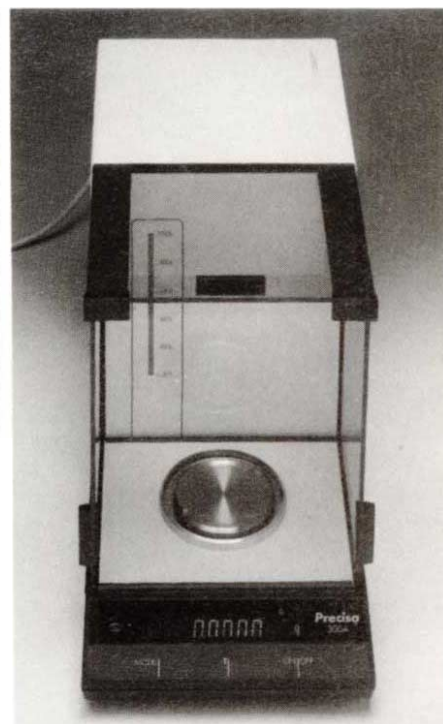
of electrodes. The Infit 767 is available in two versions: model 767-12 is 12 inches in diameter and has an insertion length of 200 mm, and model 767-19 is 19 inches in diameter with a 200–300 mm insertion length.

Treff AG has incorporated its years of experience in precision injection moulding into the design of its **microlitre pipettes** (*Reader Service No. 104*). Treff's pipettes are available in long or short styles to suit the application, and are made to fit the human hand comfortably. For use with its pipettes, Treff also offers an autoclavable plastic box containing a tip rack for storing pipette tips.



Treff's approach to small volumes.

The 290 series of **analytical and semi-micro balances** from Precisa come with a built-in calibration weight which automatically calibrates the balance to within 0.1 mg every 24 hours (*Reader Service No. 105*). The five analytical balances in the series have a weight capacity of 300 mg, reported in 0.1-mg increments. The semi-micro balance has two weight ranges: 0.01 mg to 40 g, and 0.1 g to 200 g. Precisa's balances can be programmed through set-up, main and mode menus using only the tare, on/off and mode keys. Control, programming and command functions are entered directly into



Precisa's 300A: self-calibration is a feature.

scrolling menus on the display, and can be stored for later use. Various program functions and applications such as calculating in other units, counting, and calculating percentages are standard. A capacity and residual tare display with percentage and overload indicator are incorporated into the draught-guard at eye level. An RS232 port is also included for connecting the balances to a computer.

Socorex has a multichannel **adjustable micropipette** for use with detachable polypropylene tips (*Reader Service No. 106*). The micropipette can deliver simultaneously eight or twelve samples of between 5 µl and 200 µl by means of two pipette sizes. For use with 96-well micro-

ADVERTISEMENTS

Manipulators with piezos

Piezoelectric Manipulator PM 20 N

for fast stabbing movement. Maximum travel 20 μm . Step lengths up to 18 μm . For all intracellular manipulations. With booster up to 250 $\mu\text{m}/\text{msec}$ = 250 mm/sec! Vibrator frequency 0–100 Hz.

Step Motor Manipulator MF 500

for positioning microelectrodes in biological tissue. Precise step movement with minimal lateral vibration and no backlash at speeds up to 1 $\mu\text{m}/\text{msec}$. Total travel 45 mm!

Hybrid Piezoelectric Manipulator PM 20 H

Ideally suited for all cell manipulations. Step lengths between 0.5 and 9 μm . This is a combination of the piezoelectric device PM 20 N (see above) with a manipulator driven by motors along all three axes.

Advantage: FAST by piezo technique, LARGE TRAVEL by motor drive on three axes. Coarse adjustment by hand. Available with vibrator up to 100 Hz.

With booster: step speeds up to 250 $\mu\text{m}/\text{msec}$!

Vibrator with variable frequency available.

NEW VERSION of the apparatus position (4): PM 10 H

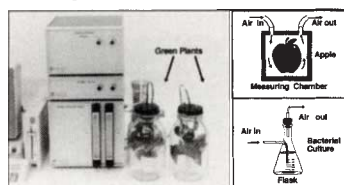
Simplified control, no digital display of travel, somewhat lower precision. Has piezoelectric translator like the one described in position (2), maximum travel 10 μm . **Extremely economical version for manipulations of and within cells!**

BACHOFER GmbH — Reutlingen D 7410
P.O.B. 7089 — W. Germany Fax 7121-54000

Reader Service No.47

O₂/CO₂ micro-Respirometer

Measures gaseous O₂/CO₂ uptake or production with 0.2 $\mu\text{l}/\text{h}$ accuracy
Far superior to the Warburg Apparatus



It can be used for measuring O₂/CO₂ production/consumption of insects, fish, small animals, plants, fruits, vegetables, bacterial and tissue cultures, fermentation, oxidation of metals and plastics etc. The results are printed periodically by IBM-PC/XT/AT.

Equipment is also available to measure oxygen consumption in larger animals (mice, rats, pigs, horses etc.).



COLUMBUS INSTRUMENTS

P.O. BOX 44049 COLUMBUS, OHIO 43204 U.S.A.
PHONE: (614) 488-6176 FAX: (614) 276-0529
TLX: 246514

Reader Service No.24

nature
is available in
microform
from University
Microfilms
International.

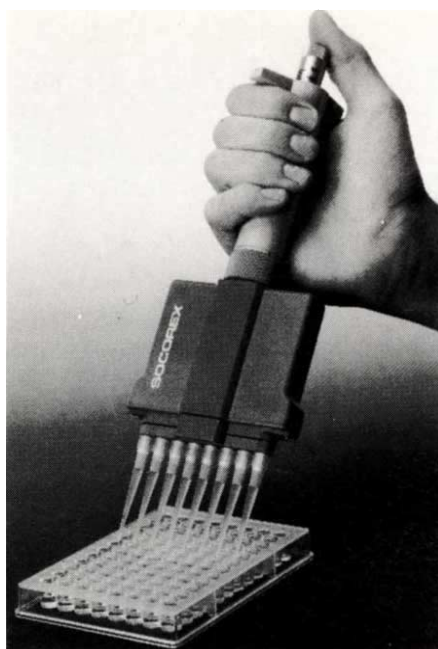
☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

Call toll-free 800-521-3044. Or mail inquiry to:
University Microfilms International, 300 North
Zeeb Road, Ann Arbor, MI 48106.

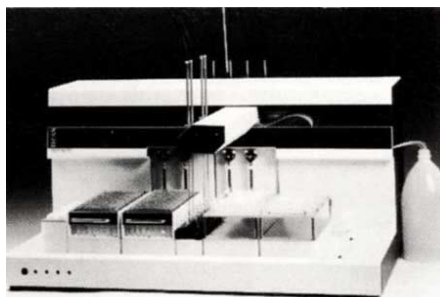


Socorex 851 aims at standard 96-well microplates.

plates, the pipette can be adjusted to 9 mm spacing. The volume is set using a Swiss built-in micrometer, and the operator uses an overshoot movement to flush the tips. An integrated tip ejector is provided to allow the tips to be changed quickly. Each instrument is individually calibrated and comes with a control certificate stating it to be accurate and precise to less than one per cent.

Assays

Tecan AG has launched a new **robotic sample processor** which allows the independent movement of the sample tips along the X- and Y-axes (*Reader Service No. 107*). Tecan says its variable-geometry model RSP 8000 is capable of processing more than 2,000 samples per hour. Each sample tip has an independent liquid level detector and is coated with Teflon for easy cleaning and protection



Swiss robotics from Tecan.

against contamination between wells. The tips can simultaneously access test tubes, reagent bottles and microwells spaced 8–280 mm apart. Tecan's model RSP 8000 also has an optional positive sample identification feature.

Anawa Laboratories, located in the Swiss Bioscience Park near Zurich, offers

a range of **contract services** for the life sciences (*Reader Service No. 108*). Drawing on expertise in the fields of biochemistry, immunology, cell biology, radiochemistry, organic chemistry and analytical chemistry, Anawa assists customers in product development and production. Anawa's services include: analysis of drugs and metabolites, drug toxicity testing in cell culture, pharmacokinetic investigations and clinical trials, drug screening, isolation and labelling of biomolecules, immunotoxicology testing and immunodiagnostics development. The company has special expertise in analytical assay development, bioassay development, and immuno- and receptor-assay development. Anawa has an array of technical equipment at its disposal, including flow cytometry instruments, atomic adsorption spectrometers, laser nephelometers, electron microscopes and nuclear magnetic resonance spectrometers.

Bioreba AG started out in 1980 selling **plant disease ELISA tests** for the detection of viruses in potatoes, grapes, barley and onions (*Reader Service No. 109*). The company has since expanded its offerings to include reagents, ELISA plates, equipment for sample preparation and liquid handling, sap extractors and plant borers, and plant testing services. The company carries out research at its own laboratories in Basel, and also works in cooperation with the Federal Agricultural Research Station of Chanigns in Nyon. New disease tests from the company include ELISAs for apple mosaic virus, prune dwarf virus, plum pox virus, raspberry ringspot virus, *Cymbidium* mosaic virus, and beet mild yellows virus.

Virion, located in Cham, produces reagents for the **serological diagnosis** of infectious diseases (*Reader Service No. 110*). All of the critical reagents, such as antigens and control sera, are lyophilized to extend their shelf lives to over 10 years, says Virion. Virion has over 50 antigens for complement fixation, including Influenza and varicella-zoster viruses, adenovirus, cytomegalovirus, *N. gonorrhoeae*, and mycoplasmas. Special antigens include those for coxsackie B 1-6 viruses, tick-borne encephalitis virus, *Camphylobacter jejuni*, five serotypes of *Leptospira*, and *Trichomonas vaginalis*. Virion offers do-it-yourself enzyme immunoassays, where the microtest plates or strips are coated by the user to save money. Immuno-fluorescence antibody tests offered by Virion include those for Epstein-Barr virus, *Borrelia burgdorferi* and *Legionella pneumophila*, sold as antigen slides or complete kits. For the purposes of retrovirus screening, Virion sells Western blot test kits for HIV-1 and slides for HTLV-1 antibodies.

Tools for analysis

Metrohm, based in Herisau, 60 km east of Zurich, has been involved with electrochemical analyses for over 40 years — the first Metrohm pH meter was produced in 1947. Recent introductions from Metrohm include new accessories for the model 690 manually injected **ion chromatograph** (*Reader Service No. 111*). The model 697 isocratic pump and the model 698 sample changer provide automatic operation for routine determination



Chromatograph, sampler and pump. All from Metrohm.

of anions and cations. The sample changer can be fully integrated with the ion chromatograph so that the autozero function is triggered prior to each new injection, and the baseline allowed to stabilize for the subsequent chromatogram. The sample changer can also be used with an ion chromatograph fitted with two columns, for the simultaneous determination of anions and cations. Using the twin system, Metrohm says ions such as chloride, nitrate, sulphate, sodium, potassium, ammonium, calcium and magnesium can be determined in 16–20 minutes. Metrohm has two columns under its Super-Sep label for the analysis of anions and cations. The cation Super-Sep column is designed to separate divalent and monovalent cations, eliminating the use of costly gradient elution systems.

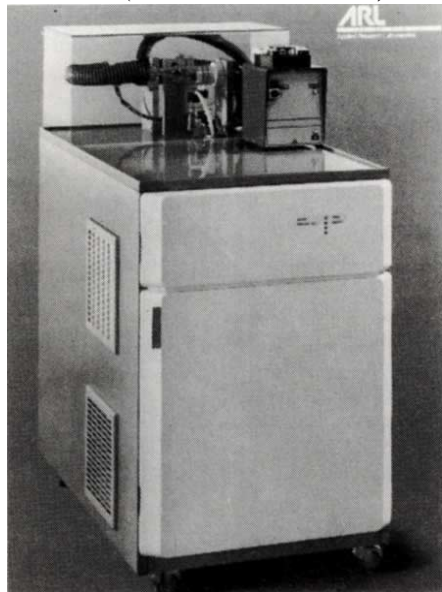
Cambridge Lasers is now distributing in the UK products made by the Swiss company Contraves, a manufacturer of electronic medical laboratory instruments. Contraves' Autolyzer AL801 is an absolute volumetric **blood analyser** system designed for medium- to large-sized haematology laboratories and emergency



For haematology, the Contraves Autolyzer.

laboratories (*Reader Service No. 112*). The Autolyzer has an automatic threshold setting which enables it to adapt automatically to the sizes of any cell population, allowing it to be used for other applications as well. Samples are presented in an EDTA tube to one of two suction cannulae: one for venous blood, which requires a sample of 250 μ l, and the other for pre-diluted capillary blood of 25 μ l in volume. Eight measurements, including cell volumes, haemoglobin content, and platelet evaluation, are automatically printed out, and another 12 can be requested by the user. Results can be printed out, displayed on the monitor, stored, or transferred to an external data processing system. Contraves says the Autolyzer can process up to 75 patient samples per hour. Optional equipment for the Autolyzer include a ticket printer, a fluid controller and an automatic sampler.

Applied Research Laboratories, located in Eclubens, develops products for inductively coupled plasma spectroscopy. ARL has a new **ultrasonic nebulizer** intended to solve the sample loss problems encountered when using pneumatic nebulizers (*Reader Service No. 113*). The



Low detection limits are the goal of the ARL nebulizer.

ultrasonic nebulizer puts more of the sample into the measuring zone to improve detection limits by a factor of 10–20, says the company. The unit is a complete system, containing electronics, desolvation unit, and heating and cooling controls. The instrument's transducer has a planned lifetime of over 500 hours. The nebulizer can be used with samples which occur naturally in solution, or which can be put into solution. It works by pumping the sample via a peristaltic pump onto the face of a piezoelectric transducer which is energized at 1.4 MHz by a radio frequency generator. Physical movement of the transducer at that frequency produces

standing waves in the liquid sample leading to the production of uniformly small droplets. The resulting vapour cloud is swept with argon into a heated chamber, which separates the solvent from the sample through evaporation and condensation. The dry aerosol then is drawn into the ICP torch for analysis.

Biochemicals

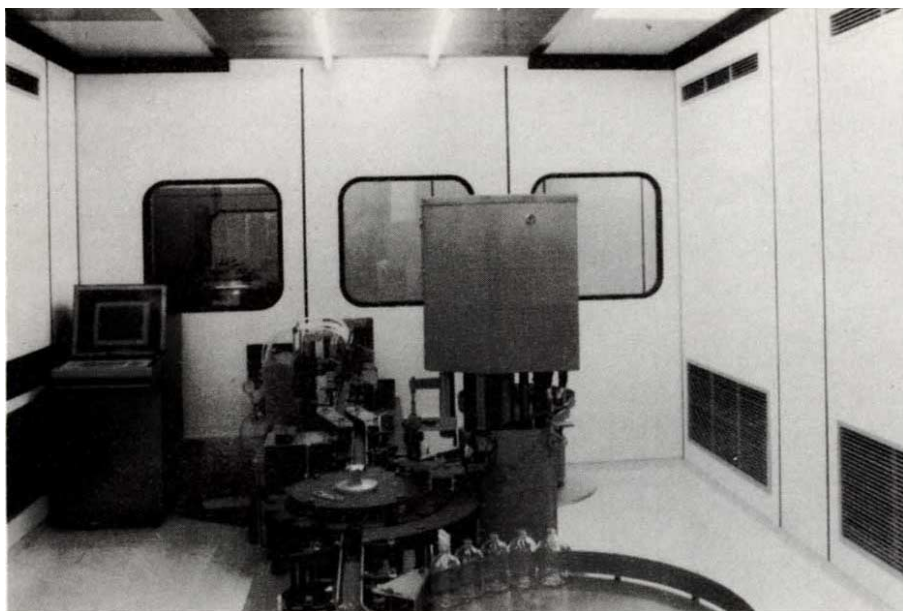
Protogen AG, based in L  ufelfingen near Basel, isolates **proteins** such as enzymes, inhibitors, growth factors and antibodies, for use in research or industrial applications (*Reader Service No. 114*). The company will also perform the large-scale isolation of proteins from a production stream on a contract basis.

Fluka's Chemika-BioChemika division offers a host of **organic compounds**, including several new ones for amino acid synthesis and the testing of nuclear magnetic resonance equipment (*Reader Service No. 115*). Fluka sells both enantiomers of 1-BOC-2-*tert*-butyl-3-methyl-4-imidazolidinone and 1-Z-2-*tert*-butyl-3-methyl-4-imidazolidinone, which are used for the synthesis of D-amino acids by alkylation or hydroxyalkylation with aldehydes. Both enantiomers of 1-phenyl-2,2,2-trifluoroethanol — chiral solvating agents — are also available for the determination of optical purities by NMR. For tetrafluorosuccinylations such as Friedel-Crafts reactions, Fluka has tetrafluorosuccinic anhydride.

Pentapharm, with operations in both Switzerland and Brazil, specializes in enzymes isolated from **snake venoms** (*Reader Service No. 116*). Snake venom enzymes offered by the company include fibrinogenase, prothrombinase, factor X-activator, protein C-activator and thrombocytin. Pentapharm is also able to isolate and purify on a large scale a variety of carbohydrates, lipids, peptides and proteins from animal tissues and body fluids, for use in pharmaceuticals, diagnostics, cosmetics and biotechnical production systems. Liver, spleen, heart, suprarenal gland and pituitary gland extracts and fractions are available, as are purified active substances such as aprotinin, collagen, glycogen, hyaluronic acid, orgo-tein, oxytocin and thymosin.

Industrial equipment

Nova Swiss has a **computer-controlled HPE plant** for the solid/liquid extraction and two-step fractional separation of industrial products (*Reader Service No. 117*). The plant is available in laboratory, pilot, semicommercial and full-scale sizes, and can be run in both automatic and manual modes. The temperature, time and extraction and separation pressures are controlled by computer; the flow rate of the mobile phase is metred by a specially



A Sulzer clean room set up for IgG production at Baxter Travenol in Lessines.

designed pump. The computer screen of the HPE plant displays a flowsheet outlining the actual process configuration, the status of all MD-valves, and the actual values of the 12 measured parameters (three pressure recordings, eight temperature readings, and one flow rate). Trends for any two selected process parameters can be displayed simultaneously over time.

Stawag has a **rotary crossflow filter** designed especially for the biotechnology industry (*Reader Service No. 118*). Stawag's Funda Type C filter has a rotating, self-cleaning filter element which the company says allows for efficient cleaning and washing, resulting in longer filtration cycles. Stawag says prototype tests using both whole and fragmented *Escherichia coli* demonstrated the Funda Type C filter's high filtration rate per unit area, efficient washing and drying of concentrated sludge, fast removal of dry sludge, and production of a high quality filtrate. The Funda Type C filters come in internal diameters of 100, 200 and 400 mm, with filter surfaces ranging from 0.06 m² to 3.0 m².

Sulzer installs custom **clean rooms** for pharmaceutical production and other sterile applications (*Reader Service No. 119*). Sulzer recently installed a clean room in Baxter Travenol's immunoglobulin production facility in Belgium. Baxter Travenol's clean room consists of several separate zones, constructed to different clean room classifications, ranging from 100 to 100,000 according to US federal standards. The air overpressure is maintained at values of between 60 and 15 Pa in the individual rooms, and air volume regulators are fitted to the fans. Each zone has its own air treatment plant, and room temperature is kept constant at 19 °C and

room humidity at 50 per cent. Lighting is provided by tubular fluorescent lamps with diffusers specially designed to influence air flow in a manner that prevents cross-contamination.

Fryma has a **ball-mill** with applications from the processing of biotechnology products to making chocolate (*Reader Service No. 120*). Fryma's CoBall-Mill 100 is a stirrer-bead mill for large throughputs in the finest grinding range. The CoBall-Mill 100 has a 160 kW capacity, and processes between 3,000 and 6,000 kg of product per hour, says Fryma. The ball-mill has an electronic monitoring system that regulates power input, mill shaft r.p.m., product input, heating and cooling, media input and temperature. Fryma says the product's greatest advantage is cost-savings in grinder beads: the company says the CoBall-Mill 100 requires 75 per cent less beads than conventional stirrer-bead mills in the processing of one

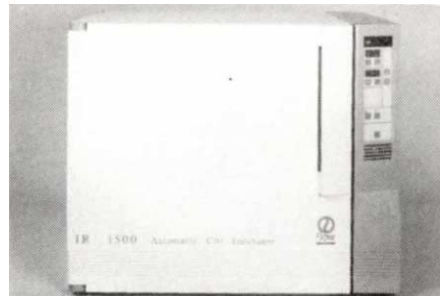


One of the Fryma range of ball mills.

ton of product. The horizontal vessel allows the for easy loading and unloading.

Conductive environments

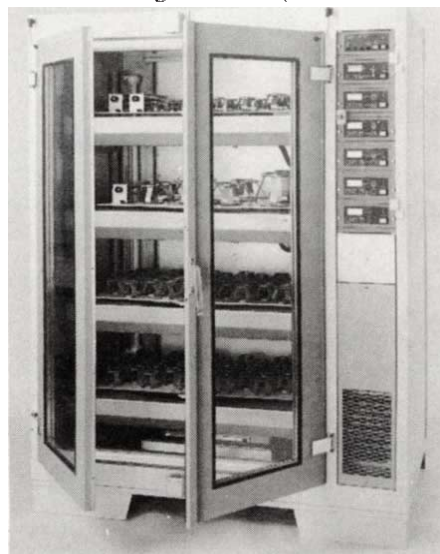
Flow Laboratories has an **automatic CO₂ incubator** that fits on the benchtop (*Reader Service No. 121*). Flow's Incubator 1500 is a cell culture incubator with a volume of 150 litres. The seam-free interior



Go with the Flow IR 1500?

chamber is made of stainless steel, and the double-glazed inner door is designed to minimize condensation and heat loss. Flow says the internal fan is constructed to restabilize quickly conditions within the chamber after the inner door has been opened temporarily.

The Climo-Shake **environmental shaker** from Adolf Kühner AG has a stainless steel cabinet and four independently controlled shaking machines (*Reader Service*



Climo-Shake can eliminate the need for temperature/humidity controlled rooms.

No. 122). Linear motors drive the shakers with electronic commutation and magnetic power transfer. The Climo-Shake can be programmed for a variety of temperatures, humidity levels and shaking rates through its digital microprocessor. Built-in refrigeration allows operation at 15 °C below ambient. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal.